

GM foods debate needs a recipe for restoring trust

The soundest possible science must underlie any effort to regulate genetically modified foods. But regulations must also acknowledge uncertainty — and incorporate trust in the judgement of the consumer.

Should the rules for labelling genetically modified (GM) food be determined solely by technical considerations, based on scientific 'facts'? Or should broader public concern be taken into account? The issue will be high on the agenda of the Codex Alimentarius Commission, the archaically named body that sets international food standards, when it meets next week to discuss labelling principles. It is not just theoretical. If the current disagreement between the United States and Europe on this issue were eventually to provoke a formal dispute over trade practices to the World Trade Organization (see page 641), then, under new WTO rules, the Codex's ruling would be critical in determining the outcome.

The dispute itself can be reduced to relatively simple terms. Members of the European Union, conscious of public pressure over GM foods, argue that any food containing detectable GM ingredients should be labelled. In contrast, the United States, apparently concerned at the extra costs of disaggregating basic food components according to whether they have been produced with such techniques, and worried that a GM label could trigger what it considers unnecessary worries about safety, says that labelling should only be required if a product is substantially different from food already available.

Exaggerations

Two points about the scientific aspects of this dispute must be made immediately. The first is that much of the recent outcry about the potential dangers of such foods, particularly in Britain, has been based on exaggerated claims, often invoked deliberately by mass-market (and occasionally more responsible) newspapers as little more than a device to increase sales. There is as yet no substantial evidence that GM foods are inherently more dangerous than conventional foods just because they have been produced using novel techniques.

The second point, however, as a series of articles in this week's issue demonstrates, is that a number of uncertainties about the full effects of such foods remain on the table (see Briefing, pages 651–656). In the case of human health, these include potential allergenic reactions to genetic changes that are not completely understood. As for the environmental impacts, many scientists feel that widely quoted 'hazards', such as the potential spread of herbicide-resistant 'superweeds', have been overemphasized by critics. But there is a broader consensus that the potential ecological disturbance caused by a growing dependence on GM crops by modern farmers could be significant.

Some argue that neither set of concerns is weighty enough to warrant the imposition of separate regulatory structures for GM foods on top of those that already exist for conventional foodstuffs. Technically, perhaps, the argument is correct. And the food industry would certainly like that view to prevail. But it fails to take into account an additional factor that must be incorporated into any regulatory system if it is to achieve its goal — the need for public acceptance.

The public is right to be concerned about the potential — and novel — hazards of modern food-production techniques. In Europe, at least, the recent epidemic of bovine spongiform encephalopathy

(BSE) among cattle, apparently the end-result of more cost-effective feeding processes introduced in the early 1980s, is a dramatic illustration of unanticipated dangers. And in both Europe and the United States, the steady reduction of biodiversity remains a silent witness to the potential of modern agriculture to inflict damage on the environment and the wildlife it supports.

Benefits

There is, of course, an upside as well. Certain consumer demands will become easier, and possibly cheaper, to meet with the new GM crops (as they are already in the case of soy protein). Agriculture will undoubtedly become more efficient as genetic modification gives farmers greater control over the range of crops that can be grown cost-effectively; in some cases, this could well allow farming communities to survive that might otherwise disappear. And it can certainly be argued that, at least in the short term, herbicide-resistant crops may lead to a reduction in the amount of herbicide used.

But neither the upside (as the industry would like) nor the downside (as environmentalists argue) should dominate the debate. A rational strategy requires an approach that respects and embraces both sets of arguments. There is no simple, institutional formula for achieving this. But some principles can be suggested.

First, both sides should accept the need to ensure that the regulation of GM foods — including the conditions under which they are marketed — is based on the soundest possible science. Basing regulations on scientific conclusions that later turn out to be false is in no one's interests; hence the need for continued research, whether this involves monitoring for long-term health effects, or field trials to study the impact of GM crops on local biodiversity. Disrupting such trials only serves the interests of those who seek gratuitous publicity.

Second, both sides should acknowledge the current limits to scientific certainty. The failure to 'prove' scientifically that a new food is dangerous is not the same as to have 'proved' that it is safe — a lesson learnt from the BSE affair. The best that research can do is to narrow the limits on uncertainties, not eradicate them.

The third need is to find ways of facilitating public access to credible scientific information — and of communicating in a responsible form both its significance and its limitations. Too much such information is tainted by its deliberate use by both sides in what can be little more than a propaganda war. As some delegates at the recent Biovision conference in Lyons pointed out, the need for 'honest brokers' is of paramount importance (see *Nature* 398, 360; 1999).

Finally, broad public concerns, however 'irrational' they may appear to some, must be taken into account in food safety regulations if they are to maintain their credibility. Industry complains that the public has lost trust in its scientific experts, but it will only make matters worse by declaring its own loss of trust in the judgement of the consumer. If labelling all foods produced by GM techniques, as many argue, turns out to be a necessary step in regaining trust on both sides, it could be a small price to pay. □



Europe and US in confrontation over GM food labelling criteria

[LONDON] The United States and the European Union are poised for a major confrontation over the regulation of genetic modification (GM) technology during international negotiations next week on the labelling of GM foods.

At a meeting of the United Nations food standards body, the Codex Alimentarius Commission, the European Union (EU) is expected to argue in support of a draft proposal that says all foods with a detectable GM ingredient should carry a mandatory label.

This proposal includes the labelling of GM foods whose composition, nutritional value and intended use are different from existing foods, even if a GM protein or DNA is no longer detectable in the finished product.

But the United States, along with Canada, Mexico, Peru and Brazil, is likely to insist that labelling should be compulsory only for foods with detectable GM ingredients, and which "differ significantly from a corresponding existing food". These countries oppose comprehensive labelling because there is no evidence of specific health hazards.

A likely focus of debate at the meeting will be the issue of 'substantial equivalence'—the concept that GM food can be considered equivalent to non-GM food providing there are no changes to its composition, nutrition-



Battle field: opposition to GM crops by environmental groups has stimulated public fears about their safety and demands that all GM food should be clearly labelled. But the United States, backed by several countries, continues to argue that there is no scientific basis for this.

ADRIAN ARBIB/STILL

al value and intended use (see Briefing, page 652). Countries such as India, Norway and Denmark, together with consumer groups, argue that substantial equivalence has yet to be proven scientifically.

These countries have joined Consumers International (CI), a worldwide federation of 246 consumer groups, in calling for all foods produced with GM techniques to be labelled as such, regardless of the presence of GM ingredients in the final product.

The disagreement hinges on whether consumers have a right to know whether a product has been manufactured using GM

technology. The United States is opposed to this, arguing that such labelling implies that GM technology is inherently unsafe, a conclusion for which it says there is no evidence.

The United States says that such labelling will be costly and complicated where food is manufactured from ingredients taken from a number of sources. It does not see a need to label food containing novel genes if its nutritional content, for example, is unchanged.

Stuart Eizenstat, US under-secretary of state for economic, business and agricultural affairs, wrote in the *Financial Times* last week that the United States will not accept "labelling that is misleading and [which] implies GM products are somehow dangerous or of lesser quality, when scientific evidence, testing and approvals show no risk to human health".

In a strongly worded attack on the EU, Eizenstat alleged that its procedures for testing and approving biotechnology products were "not transparent, not predictable, and not based on scientific principles". He added that EU calls for more research on the safety of GM products were a smokescreen "to justify keeping its trade restrictions in place".

But CI says that labelling is not just an issue of health and safety. Julian Edwards, its director-general, said in a statement that "there is no difference in the 'safety' of *halal* or organic foods, for example, but they are, nevertheless, labelled to enable consumer choice".

CI says that surveys from many countries indicate widespread public support for comprehensive labelling of GM foods. For example, 92 per cent of respondents to a survey by the UK Consumers' Association wanted GM food to be labelled, regardless of the presence of a GM ingredient in the final product.

The outcome of the debate could be significant if the United States makes a complaint about European procedures to the World Trade Organization. Ehsan Masood

US sets up 'round-table' talks with scientists

[WASHINGTON] The US State Department has invited scientists to organize a round-table conference on genetically modified (GM) foods. The meeting would be the first of a series of discussions with researchers intended to strengthen officials' grasp of key technical issues.

Frank Loy, under-secretary of state for global affairs, also says the department should appoint a science adviser with direct access to the secretary of state, Madeleine Albright.

Speaking last week at the annual science policy symposium of the American Association for the Advancement of Science (AAAS) in Washington, Loy announced a five-point plan to address criticism of how

his department handles scientific and technical issues. "We're very sensitive to your concerns," he told those at the meeting.

Loy said the new science adviser should serve the secretary of state and be supported by a small number of staff. He hopes to have the adviser in place by next autumn.

The adviser should organize "round-table discussions" involving "recognized experts on a particular issue", said Loy.

Rather than waiting for the adviser's appointment, he proposed that the department, the National Academy of Sciences and the AAAS should "work together right now to organize the first of these discussions". The topic would be GM

organisms, "particularly genetically modified agricultural products".

He said that trade in such products "will pose major issues for US policy-makers in the years to come".

Loy proposed more science training for department staff, initiatives to bring more scientifically qualified staff into the department, and greater effort to "use science as a tool for diplomacy".

Robert Stern, a consultant who chairs the industrial science section at AAAS and has worked on the State Department issue, says that Loy's offer to consult with the community through the AAAS represents a challenge for the organization. "Now we've got to organize a response." Colin Macilwain

Europe will fly animals on space station

NASA

[WASHINGTON] European researchers will be allowed to propose experiments using laboratory animals for the International Space Station, following an 11 to 1 vote last month by European Space Agency (ESA) partners in the project.

The move follows acceptance by Germany, the largest European contributor to the space station, that animals are essential for certain experiments in space. Work with higher mammals such as dogs and primates will still be forbidden. But researchers will be able to experiment with rodents, which ESA had prohibited in the past.

Many European scientists had seen the space agency's unwritten rule against rat and mice studies as a competitive disadvantage when proposing experiments in developmental biology, muscle deterioration and other areas of space research (see *Nature* 391, 733; 1998). The other major station partners — the United States, Japan, Canada and Russia — have no such prohibition.

Germany's long opposition to animal studies has been the main factor behind ESA policy. But the agency's life science working group, which gives advice on space biology research, lobbied over the past two years to have the ban removed.

"The real work was done country by country," says Didier Schmitt, head of life science research at ESA's ESTEC research centre in the Netherlands.

The German space agency DLR concluded in February that "animal experiments are unavoidable" for certain kinds of space research, says Günter Ruyters, head of the agency's life and microgravity sciences programme.

The German science ministry agreed, but



Learning experience: a Neurolab crew member is shown how to operate one of its rodent habitats.

stipulated that such experiments should be conducted only when there is no alternative, and when the scientific return is high. That cleared the way for Germany to vote 'yes' at a meeting last month of ESA space station partners to consider allowing animal studies.

Sweden cast the only 'no' vote, out of concern over possible political damage to the space station project. Per Magnusson of the Swedish National Space Board says European partners are taking a risk by approving experiments that a large fraction of the population opposes. Ruyters acknowledges that animal experiments are still "a very touchy issue" in Germany, where an already strong animal rights movement is gaining ground (see *Nature* 397, 461; 1999).

ESA's Schmitt says there is now an "urgent need" for an international committee to set ethical guidelines for animal experiments in space. That topic will be on the agenda at an international space life science working group meeting in Italy next week, which will also include a workshop on habitat design for rodents in microgravity.

Biologists learned hard lessons about how not to design such a habitat after dozens of newborn rats — more than half of those on board — died on last year's Neurolab Space Shuttle mission (see *Nature* 393, 4; 1998).

Before the flight, a plan to build a special cage for Neurolab was scrapped, and engineers modified an earlier design that had successfully housed adult rats, but not newborns. Problems with the young rats moving around on smooth (as opposed to mesh) surfaces in the cage in weightlessness may have contributed to the high death rate.

Visibility into the cage was also limited, which made it difficult to monitor the animals. A panel set up by the US space agency NASA and the National Institutes of Health to consider developmental biology research in space recommends that communications be improved among cage designers, scientists, astronauts and NASA managers to avoid such mishaps. And animals may need to be monitored more closely in orbit — either by astronauts or by ground investigators through a video system — to ensure their health.

But such problems should be surmountable, according to the panel. Despite the loss of the Neurolab rats, the panel concludes that scientific results from the flight clearly demonstrate that complex animal studies are possible in space.

Tony Reichhardt

UK's royal societies oppose new research council for Scotland

[LONDON] Research in Scottish institutions should continue to be funded through the UK research councils once Scotland gets its own parliament and executive after elections next month, recommends a report by the Royal Societies of London and Edinburgh.

The report, published today (22 April), argues that any fragmentation of funding would harm Scottish science, for example if Scottish researchers were prevented from competing for UK-wide funds.

It adds that a larger overall science base is better equipped than a smaller one to weather changes to national research priorities. And it argues that a separate Scottish science base could harm the government's plans to develop Britain's knowledge industries, potentially impeding the flow of economic benefits from basic research to Scotland.

"A large-scale research system has a greater capacity to maintain research diversity, and thereby the flexibility to pursue new directions," say the two societies. "It is therefore important that devolution does not lead to fragmentation of basic science, engineering and technology (SET) in the United Kingdom, and that Scotland remains integrated within the UK system as part of the European SET base."

These conclusions are unlikely to be welcomed by the pro-independence Scottish National Party (SNP), which is running a close second to the Labour Party in opinion polls for the 6 May elections. The SNP would like the new parliament to be as powerful as possible.

The Labour government, on the other hand, may be more supportive. Labour has placed education and the knowledge

economy at the heart of its Scottish election manifesto, promising to invest £100 million (US\$161 million) in research infrastructure over the next three years.

The report's other recommendations include the appointment of a "senior minister" for science, and the creation of a science policy advisory board for the Scottish executive. Chaired by a senior scientist, this board would comprise representatives from research and industry.

Another senior scientist should be given responsibility for science advice to ministers, for the implementation of science policy, and for liaison with British science bodies, says the report.

The report adds that members of the Scottish parliament in Edinburgh will need their own source of independent scientific advice.

Ehsan Masood

Transplant panel to play 'honest broker'

[PARIS] A small group of academics and health officials will meet in Canada in June to discuss new ways of generating public debate on the risks and benefits of the clinical application of xenotransplantation — the transplanting of animal cells, tissues and organs.

The meeting's organizers hope that this issue will serve as a testbed for the creation of an international think-tank covering ethics, science and governance.

The meeting at Meech Lake, Quebec, is the brainchild of Fritz Bach, a xenotransplant scientist at Harvard Medical School. Given the risk that xenotransplants may create pandemics, Bach has argued for a moratorium on clinical trials until there has been a public debate at the international level (see *Nature* 391, 320–325 & 326; 1998).

The meeting will be co-chaired by Strachan Donnelly of the Hastings Center in New York State and Farkhonda Hassan, a member of the Shoura Assembly (senate) of Egypt and professor of the science department at the University of Cairo. It will bring together more than a dozen xenotransplant experts, ethicists and non-governmental organizations, such as Pugwash.

Elizabeth MacGregor, former coordinator of the government's National Biotechnology Advisory Committee, has been responsible, along with Bach, for organizing the meeting.

"The meeting is expected to call for an international panel on xenotransplantation and the convening of a broad public consultation in liaison with national and international organizations to produce a detailed case study and international guidelines" says Bach.

"The choice of panel members will be crucial to its validity" he adds, arguing that it should include representatives of the various points of views on xenotransplantation, including the extreme views.

Xenotransplantation will serve as a test case for the organizers' more ambitious idea of creating a broader international think-tank. Discussions are at an early stage, but it has already attracted the interest of some prominent academics, including Keith Bezanon, director of the Institute of Advanced Studies at the University of Sussex, and Geoffrey Oldham, former director of the university's Science Policy Research Unit.

No name for the body has yet been decided, although it would be along the lines of The International Council on Science for Policy. The think-tank would identify other areas of technology where the risk-benefit equation is difficult to resolve, and would seek ways to better engage public debate and provide impartial information through setting up a series of working groups.

Bach sees the think-tank as a body free from political pressures, commercial inter-

ests and excessive scientific enthusiasm.

Abdullah Daar, from the Sultan Qaboos University in Oman, who will attend the Meech Lake meeting, believes there is a need to create such a body to break the excessively uniform thinking in many regulatory agencies and institutional ethics bodies. "National and professional organizations tend to mirror their perception of what the public ought to accept, rather than saying these things need to be thrashed out first," he says.

For the moment, the main outstanding questions are "the council's scope, how best to consult the public and who should choose its members" says Oldham. "We share a concern to find international approaches which genuinely involve developing countries, and a belief that these approaches must involve consultation with relevant stakeholders, including the public".

"Any mechanism that is devised must be complementary to what others are doing.

Hence our current concern is in identifying an appropriate niche." Oldham emphasizes that if the idea is to work it will require "major efforts to reach decision makers and to try to ensure that action follows".

Margaret Somerville, from the McGill Centre for Medicine, Ethics and Law at McGill University, Montreal, welcomes the proposed think-tank. Existing ethics committees tend to be made up of representatives of various constituencies, she argues, so they often act as political bodies, with consensus being negotiated on the basis of "what do we do here to make a compromise to get the best deal for what we want to promote".

Somerville says there is a need for transnational structures, with no vested interests, that can put complex issues on the table, provide impartial information on their various aspects and solicit wide input, while being free from the pressure to reach an immediate consensus.

Declan Butler

Looser ties urged for Japanese scientists

[TOKYO] Japan's trade minister, Kaoru Yosano, last week called for university scientists to be given more freedom to work in collaboration with industry. His comments were made in response to recent controversy over Sony's appointment of an economics professor from Hitotsubashi University to its board of directors.

Such external work is forbidden by civil service law, which prohibits employees of national universities from taking part in profit-making activities. But Yosano suggested that university researchers be given an exemption on the basis of "the potential benefit of collaboration between industry and the academic community".

The Ministry of International Trade and Industry (MITI) and the Ministry of Education, Science, Sports and Culture (Monbusho) have been working towards relaxing the law (see *Nature* 390, 105; 1997), but little progress has yet been made.

The delay is partly due to concern that relaxing the law may trigger corrupt relationships between industry and universities. MITI officials attribute current sensitivity to a recent bribery scandal involving a former professor from Nagoya University and a major drug company (see *Nature* 396, 205; 1998).

Hirooyoshi Hidaka, a former professor at Nagoya University Medical School, and executives from Otsuka Pharmaceutical Co. were found guilty last week over payments relating to Otsuka's drug research and development activities.

It is said that Hidaka claimed to have received payments from Otsuka for



Kaoru Yosano: wants to relax current law.

providing 'technical consultation'. But the Nagoya District Prosecutor's Office ruled that the payments were bribes, on the basis that Hidaka accepted them in return for allowing the company to use university facilities, including researchers,

for commercial research — an activity that is also prohibited by the law.

While the public has been quick to express its disapproval, many researchers consider the incident to be no more than a conventional industry-university collaboration that was considered corrupt because of a lack of clear guidelines.

"It creates a terrible dilemma, as the government is actively urging university researchers to become involved in business activities, yet such involvement is restricted by the law," says Tamon Inoue, professor of engineering at Tsukuba University and head of its Venture Business Laboratory.

Akito Arima, minister of education and director-general of the Science and Technology Agency, is a supporter of such collaboration provided that activities are restricted to basic research.

But Arima said many questions still need to be addressed, such as those concerning how much researchers should be paid if they work for a private company, and also what their status would be if the company were to go bankrupt.

Asako Saegusa

Financial doubts over future of chimp lab

[WASHINGTON] The laboratory housing the world's largest domesticated chimpanzee colony may be near financial collapse, animal rights activists are claiming.

The Coulston Foundation (TCF) of Alamogordo, New Mexico, has lost two contracts to house HIV-infected chimps and conduct AIDS vaccine research on them. The National Institutes of Health (NIH) contracts brought the facility about \$10 million over the past six years.

Many of the TCF's 620 chimpanzees were infected early in the AIDS epidemic, when scientists were seeking models for studying the disease. But they are used less now because, while infectable with HIV, they do not get sick from the disease.

The foundation denies being in financial peril, saying the loss of the NIH contracts is due to a change in the bidding process, and that it intends to bid for a new contract.

But In Defense of Animals (IDA), a Californian group that opposes the use of animals in research, claims that the lost contracts could tilt the \$10.5 million lab towards financial failure. "The loss of these NIH funds could spell the beginning of the end for TCF," says Suzanne Roy, IDA's programme director.

The group alleges that the lab's private sources of funding — about half its revenues, according to the foundation — are drying up. TCF laid off 20 employees in November.

If the lab fails it would exacerbate a difficult situation in which a surplus of US research chimps — expensive to care for and used less and less by scientists — need housing. The National Research Council recommended in 1997 that the NIH establish a \$7 million-a-year programme to care for the government's approximately 1,000 chimps (see *Nature* 388, 218; 1997). But this has not been done.



Out of contract: chimps are used less now by scientists because they do not develop AIDS.

Don McKinney, a spokesman for TCF, says that NIH did not renew the government contracts because it is moving to a competitive bidding process. "We have the right to bid on them again and we fully intend to."

McKinney declines to discuss the lab's private funding, saying that supporters are harassed by activists if they are identified. He says the activists, whom he describes as "gangsters", have been "trying to ring our death knell for several years. Obviously they weren't too accurate." An NIH spokesman says that the contracts were not cut off because of problems: "It's just that they expired."

The IDA has been investigating TCF for five years, and alleges chronic neglect of the lab's chimps. The group has been vindicated at least in part by three formal complaints

issued by the United States Department of Agriculture (USDA), charging that the lab has violated the Animal Welfare Act. The most recent was issued in February.

USDA inspectors found the lab dirty, infested and without adequate ventilation, adequate protection from extreme temperatures or adequate removal of animal and food wastes. They documented the deaths of chimpanzees from inadequate veterinary care, including failure to treat an animal for shock before surgery after it was attacked by another chimp, and failure to monitor it adequately after surgery. TCF's failure to take into account the side effects and complications of experimental drugs being tested resulted in "the unnecessary deaths" of three other chimpanzees, the USDA charges.

McKinney says the foundation "disagreed" with most of the USDA findings and has responded in documents not yet public.

Also in February, TCF was disciplined by NIH's Office for Protection from Research Risks. The office required, among other things, that TCF hire seven "fully qualified" veterinarians. McKinney says the facility, which also houses 250 monkeys, currently has 4.5 veterinarians on staff. "They are difficult to find. There are not that many veterinarians running around who have experience with non-human primates, and in particular chimps."

The National Institute of Allergy and Infectious Diseases (NIAID) confirmed last week that it had not renewed a \$640,000, one-year contract that expired on 31 March. Officials would not comment on the reasons. (Between 1993 and 1997, NIAID paid TCF \$3.3 million in contract costs.)

A six-year, \$5.3 million National Cancer Institute (NCI) contract with TCF expired on 31 December and has not been renewed. During the contract, the number of chimps at TCF supported by NCI fell from 30 to six. "The research that we had wanted to accomplish was finished," says an official. "We never had any problem with the care that the Coulston Foundation gave to our animals."

The foundation is currently receiving \$8.8 million between 1995 and 2000 from the National Center for Research Resources (NCRR), under its National Chimpanzee Breeding and Research Programme.

IDA claims that this contract will expire or be reduced next year, as the government reduces support for chimpanzee breeding. The NCRR plans to consolidate the five chimp centres it supports to two or three.

Some of TCF's chimps came from New Mexico State University in 1993, and a hundred came from New York University's Laboratory for Experimental Medicine and Surgery in Primates, which closed in 1997 (see *Nature* 390, 321; 1997). Meredith Wadman

Japan plans cDNA hunt for disease genes

[TOKYO] The University of Tokyo and 12 Japanese companies have started a joint initiative to sequence and analyse full-length human complementary DNA (cDNA) to discover patentable genes that will contribute to the understanding of diseases.

The project is intended to develop new technologies in drug development, bioinformatics and gene therapy. It will receive ¥2.5 billion (US\$208 million) in this fiscal year, which began on 1 April.

Researchers from the university's Institute of Medical Sciences will provide the companies with cDNAs from various tissues, including cancerous tissues from liver, lung and stomach, and those linked to neurological and immunological diseases.

The companies will carry out DNA base sequencing and analysis of proteins, with an

aim of discovering patentable genes for medical use and research. They plan to sequence 7,500 cDNA clones by next March.

The project involves companies such as Hitachi, Daiichi Pharmaceuticals, Kyowa Hakko and Otsuka Pharmaceutical Co., and is funded by the New Energy and Industrial Technology Development Organization, which is affiliated to the Ministry of International Trade and Industry.

The move is in line with government efforts to commercialize biotechnology research. In January, five government bodies launched a 10-year programme to expand 25-fold Japan's biotechnology market (see *Nature* 397, 554; 1999).

The programme is expected to create a market worth ¥25 trillion and to help create up to 80,000 jobs by 2010. Asako Saegusa

Spanish watchdog sees way ahead for stem-cell research

[BARCELONA] At a time when public debate in Spain appears to be moving in favour of therapeutic cloning, the country's National Commission on Assisted Reproduction has come out in support of the use of non-embryonic stem cells as a potential source of human cell cultures, tissues and organs.

Its comments are included in a series of recommendations to the government on issues related to human reproduction, such as the preservation of frozen embryos, sperm and oocytes, and the donation of embryos.

The commission is headed by Enrique Castellón, under-secretary in the Ministry of Health, and is made up of 22 specialists from ministries and scientific societies, medical organizations and lawyers. In its first annual report, it emphasizes that both ethical and legal aspects of the concept of "embryo status" provide greater support to stem-cell research than to other techniques for producing cloned embryos for research.

According to the commission, general principles about considering human beings as an end rather than a means — as well as the right to be genetically unique and not genetically programmed — constitute a serious ethical objection against reproductive cloning by nuclear transfer of somatic cells.

At present, there are no well-established research teams engaged in cloning experiments in Spain, says the report, although this situation may change soon as the centres improve their technological capabilities.

The commission emphasizes that human reproductive cloning is forbidden by law in Spain, and proposes a hardening of the penal code, which it says could be misinterpreted in its present form, to penalize such activities.

Reproductive cloning by splitting embryos has only limited potential application, and would also be problematic from an ethical point of view, says the report.

The commission acknowledges the potential advantages of using non-reproductive human cloning to obtain tissues and organs for transplantation, but argues that the "problem" of the "embryo status" discourages the development of such techniques.

As a result, says the commission, "embryonic cells must be obtained by nucleus transfer. And this means the creation of a human embryo, even though it may have a few days of life, and its subsequent destruction in the lab to obtain cell cultures."

Spain's Catholic Church has not issued any reaction to the report, but it is opposed to any form of artificial manipulation of embryos, on the grounds that "the human being has the right to be a product of the natural genetic randomization".

Xavier Bosch

Private nuclear waste plan faces critics in Australia

[SYDNEY] A cat-and-mouse game that has been taking place since 1997 over an ambitious private plan for an international repository for high-level nuclear waste in Australia has at last become the subject of public debate.

Plans for such a repository drawn up by Pangea Resources Australia were revealed last December by Friends of the Earth in London. Surrounded by accusations of secrecy and back-door dealings, the company's proposal has provoked considerable public opposition.

Pangea's leaders have now begun to talk publicly about the plan for the first time. US chairman David Pentz outlined the plans for the waste, which comes from spent fuel and dismantled weapons, at a conference in Tucson, Arizona, last month. He said that Australia had been selected as the only location to take such waste following studies of "four regions within several nations".

Pentz said that a "very rigorous set of siting criteria" favours the vast, 300–800-million-year-old sedimentary basins spanning the states of Western and South Australia, the remnants of the 'supercontinent' of Pangea.

A network of chambers 500 metres underground and covering 20 square kilometres would house sealed containers in which radioactivity could subside to natural levels over 250,000 years, according to the plan.

But some Australian geologists back contentions by environmentalists that Pangea's prediction of geological stability is unrealistic.

Pentz linked Pangea's proposal to future applications of 'Synroc', an Australian process for immobilizing radioactive material within a synthetic rock formed under high pressures and temperatures. He said he saw this as an "integral part" of preventing the proliferation of nuclear weapons.

Pangea was formed by Golder Associates, a Canadian company now based in Seattle, Washington State, and is backed by two organizations active in nuclear waste disposal, British Nuclear Fuels and Nagra of Switzerland. The company estimates the cost of constructing the site and providing transport infrastructure as US\$6 billion, and annual operating costs as \$450 million.

Pangea predicts that global production of nuclear waste will grow to more than 450,000 tonnes by 2020. It would build dedicated ships and railways to import 2,000 tonnes each year, up to a maximum of 75,000 tonnes.

Although it has no nuclear power stations of its own, Australia is a significant player in the nuclear industry, as it provides about one-fifth of the world's supply of uranium. But Nick Minchin, the industry, science and



Rough road ahead: federal ministers oppose plan to bury nuclear waste in Australia's outback.

resources minister, says user nations should be responsible for storing their own waste.

Minchin has written to Jim Voss, Pangea's senior representative in Australia, emphasizing that the government's policy of not accepting nuclear waste from other countries is "absolute and will not be changed". Minchin said the government "has no intention of considering Pangea's proposal".

Minchin says Australia is only committed to dealing with waste from its sole 'research reactor' at Lucas Heights, Sydney, and from its forthcoming replacement, which was cleared on environmental grounds by environment minister Robert Hill last month (see *Nature* 398, 454; 1999).

But Australia has yet to decide where to store the reactor's waste, accumulated over 40 years in 1,600 fuel rods. Hill has stipulated that the new reactor can proceed only when a comprehensive waste plan is approved.

Charles McCombie, a geologist and Pangea's head of science, technology and engineering, told a uranium conference in Darwin that the company had submitted a "summary project" to the federal government in February. He described deep underground disposal as a "necessary and a responsible waste management course" which he judges to be "technically feasible".

According to McCombie, the submission contains incentives of "significant economic stimulation in Australia of the order of one per cent of the gross national product, and long term employment benefits, creating as many as 70,000 jobs".

McCombie accepts that "public opposition to specific repository siting proposals is high". But he seems undeterred, and forecasts a long campaign.

Brian Anderson, president of the Australian Academy of Science, said on Australian television this week that he personally supports the Pangea proposal, and sits on its scientific review board.

Peter Pockley

CORBIS/HOWARD DAVIES

Scientists urged to raise lobbying efforts

[STRASBOURG] Members of the European Parliament (MEPs) would find it easier to secure more money for science if scientists were to lobby more actively, a meeting in Strasbourg was told last week.

Claude Desama, a Belgian MEP and former chairman of parliament's research, technological development and energy committee, said that greater input from scientists would have helped parliament in its battle last year with the European Council over the budget for the European Union's fifth Framework programme (FP5).

Desama pointed out that lobbying by scientists, scientific institutes or scientific organizations "can at best be described as discrete, at worst, barely existent".

At a meeting organized by Euroscience, a 'grass roots' association of European scientists, he said that big companies lobby hard and efficiently. In contrast, he added, "MEPs feel that they are working without a safety net — unsure of ourselves — when we have so little direct contact with scientists."

"Scientists should get together and put us under pressure," says Christof Tannert, speaker of parliament's socialist group on FP5 and life sciences. If they want to influence political reality, he says, they need to lobby alongside industry and pressure groups like Greenpeace. This would strengthen parliament's position by giving it confidence that it has the scientific community behind it.



He welcomed efforts by Euroscience to create new forums for establishing contact. Euroscience had timed last week's debate — its first such initiative — to help ensure that science is an issue in campaigns for the European parliamentary elections in June.

During the debate, several MEPs revealed their frustration with the stance taken by the European Council, whose members are made up of EU member states. "Debates over FP5 were like bargaining in a souk, with each member state discussing what its own yield on each item would be," said Desama.

Dutch MEP Elly Plouij, a member of the liberal group and research committee spokesperson, said there had been a "regrettable tendency towards renationalization" in

the council, with each country looking out for its own interests.

"When we came to debate the PHARE programme [which supports research in central and eastern Europe], some delegates argued for a lower budget so that 'there would be more left for us'," she argued.

Members of Euroscience used the meeting to argue for more money to be allocated to European-level research in the sixth Framework programme, due to begin in 2003.

As Europe moves closer to political integration, "the timing is right for a change in the distribution of European funds, so that the European Union provides a much larger proportion of total European research spending," said Peter Tindemans, a physicist who was director for research and science policy in the Netherlands ministry of education before retiring earlier this year.

The FP5 budget represents only six per cent of total European spending on research, with other European-level research laboratories, such as the European Laboratory for Particle Physics (CERN) and the European Molecular Biology Laboratory, together representing a further six per cent.

Tindemans said that these proportions should increase significantly, and that the issue should become political. Other MEPs are sympathetic to this view, aware that in the United States basic research is funded at the federal level.

Alison Abbott

Bioinformatics institute plans public database for gene expression data

[MUNICH] In a bid to speed up the exploitation of human genome sequencing efforts, the European Bioinformatics Institute (EBI) — an outstation of the European Molecular Biology Laboratory (EMBL) — is planning to launch a publicly accessible repository for DNA microarray-based gene expression data.

It hopes to create a single location where all the data on gene expression obtained from microarray technologies can be stored. But some scientists doubt whether the technology is sufficiently developed.

The EBI is organizing an international meeting later this year of representatives from laboratories that use such technologies. It will take place at the Wellcome Trust Genome Campus in Hinxton, near Cambridge, UK, where the EBI is based. It hopes to set up working groups to develop standards for

microarray-based gene expression data and analysis.

DNA microarray, or 'chip' technology, allows scientists to rapidly monitor which genes are being expressed in a particular tissue in a highly automated way. The microarrays are coated with a mixture of cDNAs (or synthetic oligonucleotides identifying particular genes) whose sequences have been identified through genome sequencing projects. These bind mRNAs, the specific messengers made by a gene when it is turned on.

Expression patterns can then be compared between healthy and diseased tissues, providing clues to the genetic complexities of diseases such as cancer.

It is not enough to know whether a gene is present in a disease, says Annemarie Poustka, senior molecular geneticist at the German Cancer Research Centre (DKFZ) in

Heidelberg. It is also necessary to know whether it is switched on, and whether it stays on through all the stages of a disease. As much data as possible needs to be pooled for this, she says.

"The microarray gene expression repository may become one of the most important databases in bioinformatics," says Alvis Brazma, a staff scientist at the EBI.

Although DNA microarray technology is in its infancy, it has already created a large amount of data, which are either held privately or scattered across the Internet. "As more laboratories acquire this technology, the amounts of large-scale gene expression data and profiles will grow rapidly," says Brazma. "This could lead to an explosion in gene expression data that may dwarf even the human sequencing projects."

But some researchers think

the move may be premature. Poustka agrees that a central public database is needed, and that the EBI is an ideal host, but says that different laboratories have not yet developed the tools to make comparisons between their data straightforward.

But the EBI, which has discussed its plans with European and US laboratories that use these technologies, is confident that the time is right to develop resources and standards.

"The database will allow us to cross-validate data obtained by different technologies, to characterize various techniques, and to establish error rates, benchmarks and gold standards," says Brazma.

EBI scientists, along with European colleagues, are applying for grants to develop the database, and hope it will become an international, not just a European, effort.

A.A.

A meeting of parallel paths for US fusion?

Colin Macilwain

The United States is realigning its nuclear fusion programme to improve the integration of research into two rival approaches to fusion power: magnetic and inertial confinement. The task is a challenging one.

[WASHINGTON] A series of top-level reviews and scientific meetings will be carried out this year to try to find ways of integrating US research in magnetic and inertial confinement fusion. The move follows the country's withdrawal last year from the main global research collaboration aimed at achieving controlled nuclear fusion.

Magnetic confinement fusion has historically been supported in the United States by the Department of Energy's Fusion Energy Sciences programme, while inertial confinement fusion is supported by the department's nuclear weapons programme.

The Fusion Energy Sciences programme was sharply cut three years ago, as Congress appeared to lose patience with the feasibility of fusion as an energy source. This was followed last summer by US withdrawal from the International Thermonuclear Experimental Reactor (ITER) project, a global collaboration to build a test-bed for magnetic fusion (see *Nature* **394**, 511–512; 1998).

But, although Congress has cooled on magnetic fusion, it has been willing to support a major new inertial confinement fusion experiment: the \$1.5 billion National Ignition Facility (NIF) at the Lawrence Livermore Laboratory in California.

At NIF, ignition is obtained by compressing a tiny pellet of hydrogen fuel with X-rays — the principle behind the hydrogen bomb. Congress supports this work because it will enable scientists to improve their understanding of nuclear weapons without testing.

Any possible energy implications are seen

as an incidental benefit. The result, however, is a reversal of the respective positions of the two programmes over the past few years. In 1995, the US spent \$360 million on the magnetic confinement fusion programme, and a little over half that amount on inertial confinement fusion. This year, it will spend \$500 million on inertial work and less than half of that on magnetic concepts.

The two programmes have many common technical needs, and the plasma physicists who lead them know each other well, having trained in the same places and attended scientific meetings together for years. But the rigid independence of the two strands of research, combined with the rapid changes in their respective fortunes, has raised questions about the balance of the overall programme as it now stands.

Last autumn, the Senate energy and water appropriations subcommittee asked the energy department to conduct a broad review of the entire fusion programme. The request produced a flurry of activity. Last month, for example, a sub-group of the Secretary of Energy's Advisory Board met in Washington to start a quick assessment of the "appropriate balance" between magnetic and inertial fusion.

The Fusion Energy Sciences Advisory Committee, chaired by John Sheffield of Oak Ridge National Laboratory in Tennessee, will report later this year on its priorities for the civilian science programme. The National Research Council is about to start an assessment of the quality of the research in the

civilian programme. And in July, a two-week workshop at Snowmass, Colorado, will bring together most of the top scientists in both programmes to try to agree priorities for research and future facilities.

All this activity is unlikely to make much difference to the inertial confinement fusion work, which will continue to be supported primarily to meet the needs of the stockpile stewardship programme for maintaining nuclear weapons. But the implications for the fusion energy sciences programme could be profound: for example, bringing together inertial and magnetic fusion experts may produce a longer wish-list of research needs than the budget can bear.

With this in mind, leaders of the two communities have been working together to draw up a medium-term plan — a 'roadmap' in the parlance of the energy department — for the Fusion Energy Sciences programme.

Mike Campbell, associate director for lasers at Lawrence Livermore, has been working with Rob Goldston, director of the Princeton Plasma Physics Laboratory in New Jersey, to thrash out details of the plan. The two have called for the community budget to be increased from \$223 million to \$300 million "in the near term", with the new money split between work to support magnetic and inertial fusion.

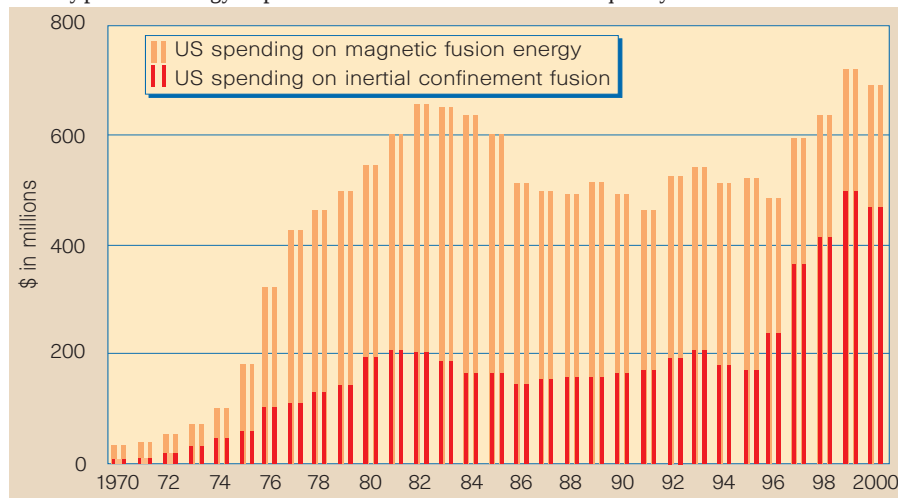
The magnetic programme wants support for experiments at existing tokamaks, and for the investigation of other confinement experiments. It also wants support for the use of overseas experimental facilities, such as JET in the United Kingdom. Goldston says: "Maybe around 2003 or 2004 it will be time to address some major next-step facilities."

In his judgement, the international partners in the ITER project will not decide to build the experimental reactor before that time, although he adds that "there might be a 'no' on ITER" by then.

The roadmap calls for research into problems that need to be solved if inertial confinement fusion is to be tapped as a power source, but which the nuclear weapons programme will not support. Both approaches to fusion, the roadmap points out, stand to benefit from additional research in materials and technology development, computer simulation and basic plasma physics.

Campbell has joined with Goldston and other magnetic fusion leaders to submit joint testimony to congressional appropriations committees, asking them to go halfway to meeting the roadmap proposals by appropriating \$260 million for the Fusion Energy Sciences programme next year.

Fusion advocates admit that such an increase is a tall order in this summer's round of appropriations, which promise to be at least as tough as previously. But without the extra money, the US magnetic fusion community will face another painful year in which to ask itself: if not ITER, what? □



Call for more research to monitor Britain's sheep for signs of BSE

[PARIS] An expert panel set up by the committee that advises the UK government's Spongiform Encephalopathy Advisory Committee last week acknowledged a "possibility" that the agent that causes bovine spongiform encephalopathy (BSE) may have passed to sheep and spread within flocks.

In its report, the panel called for a greater research effort to "assess whether, and to what extent, there may be a risk to human health". Its recommendations include a retrospective study of feeding practices to identify high-risk flocks exposed to BSE-contaminated meat and bone meal.

The report highlights the risk that BSE in sheep may have gone unnoticed, because the clinical symptoms are virtually identical to scrapie. Whereas only a handful of putative scrapie cases have been strain-typed for prion, the report calls for all future cases to be strain-typed, and for priority to be given to the development of tests to distinguish between scrapie and BSE (see *Nature* **395**, 6–7; 1998).

A comprehensive series of experiments in sheep is also called for in the report, which acknowledges that such research may be

limited initially by logistical difficulties associated with large-scale animal experiments involving prion diseases. All suspect sheep should be genotyped, it adds.

Japanese campaign to enthuse young scientists

[TOKYO] Japan's Science and Technology Agency last week launched a three-year campaign to promote the public understanding of science, intended to revive declining interest in science-related subjects among schoolchildren.

The agency plans to host science-related events during the campaign, including the opening of Science World, a centre for international research exchange. The campaign includes plans to launch a science-only television channel and a 'virtual science museum' on the Internet. The agency will also allocate a budget for improving facilities at 300 science museums across Japan.

Chile and Germany sign pledge to collaborate

[MUNICH] Germany and Chile last week agreed to increase their cooperation in science and technology and in university education. In the presence of Chile's president, Eduardo Frei, an action

programme was signed in Berlin to promote joint activities over the next three years in information technology, biotechnology, Earth sciences, and environmental and marine research.

In addition, the German and Chilean conferences of university rectors have signed a framework agreement to allow universities in both countries to cooperate in the exchange and education of students and scientists.

According to Germany's science minister, Edelgard Bulmahn, the programme is intended to put new life into a 1970 agreement on scientific cooperation. As a first step in this direction, a joint German–Chilean workshop on El Niño research is to be held in Chile in October.

Novartis plans its third laboratory in California

[SAN DIEGO] The Swiss life science company Novartis is planning its third research laboratory in La Jolla, California, where about 130 scientists will focus on infectious diseases. The laboratory will have an estimated annual operating budget of \$30 million, and will be located near the non-profitmaking Scripps Research Institute, which has a strategic alliance with the pharmaceutical company.

Novartis is also establishing the Institute for Functional Genomics, to be led by biochemist Peter G. Schultz, and the Agricultural Research Institute, headed by Steven P. Briggs, near Scripps. Novartis is said to have long-term plans for additional facilities in the region.

European Parliament extends science panel

[STRASBOURG] Members of the European Parliament have voted to expand the responsibilities of its committee for research, technological development and energy to include industry. "It is never a good idea to consider trade and research separately," said Dutch member of parliament Elly Plooij at a meeting organized by the Euroscience group in Strasbourg last week.

The move has been criticized by some scientists who fear that bringing industry and research together on to the same committee could mean that research issues will get marginalized. Christof Tannert, speaker of the parliament's socialist group on Europe's fifth Framework programme of research and on the life sciences, says that members are familiar with this concern. "But it is not our intention for research to become subordinate to industry."

France sets the scene for debate on reforms

[PARIS] Demands by many French scientists for a full national debate on research as a prerequisite for reforms appear about to be met. In February, prime minister Lionel Jospin asked a parliamentary mission to look at ways of improving links between the research agencies and the universities, and of giving greater independence to young researchers.

The plans for the mission, due to be announced in detail next week, indicate that the initial two-person mission has been expanded to include a steering committee, and will hold a series of hearings and consultations, as well as public meetings throughout France. These will culminate in a national colloquium on research in June, with recommendations being made to the prime minister in July.

UK students fight shy of physics and chemistry

[LONDON] The numbers of students applying for British university science courses in physics, chemistry and engineering have fallen substantially again this year, according to figures released last week by the Universities and Colleges Admissions

Service. Physics has seen the number of applicants drop by 10 per cent compared with applicants at a similar time last year.

Chemistry is faring almost as badly, with a drop of 9.3 per cent. In contrast, computer science is rapidly increasing in popularity, and, with a 19.7 per cent increase in its numbers of applicants, is the second most popular subject, having overtaken law for the first time. Business and management studies come top.

Gates helps fund home for MIT's computer lab

[BOSTON] Bill Gates, chief executive of Microsoft Corporation, has donated \$20 million to Massachusetts Institute of Technology's Laboratory for Computer Science. The announcement was made at a ceremony last week to honour the lab's thirty-fifth anniversary.

Gates, who delivered the keynote address, called the lab "one of the most important centres for computer research in the world". The money will help finance the construction of a new \$100 million home for the lab, being designed by the architect Frank Gehry. The building, scheduled for completion in 2003, will replace the historic Building 20, in which radar technology was developed during the Second World War.



The full text of news items on this page — and further details and background information about the Unesco/ICSU World Conference on Science, to be held in Budapest from 26 June to 2 July — can be found on *Nature's* website at <http://www.nature.com>

Nonlinear dynamics could help in battle for sustainability

[ROME] The study of nonlinear dynamics offers “unexplored possibilities” for overcoming the current “stalemate” in knowing how to face the many threats facing modern civilization. So concludes a statement issued following a study week organized last month by the Pontifical Academy of Sciences.

The statement was drawn up by Marcelo Sanchez Sornod, chancellor of the academy, and Vladimir Keilis-Borok, of the International Institute of Earthquake Prediction Theory and Mathematical Geophysics in Moscow. It has been put forward as a contribution to the deliberations of the World Conference on Science.

The authors say that the meeting extended a wide range of debates on the applications of nonlinear dynamics to critical phenomena. It also identified areas in which the resources offered by such a form of analysis “have not been fully exploited”.

In particular, the meeting suggested that one area might be the study of scenarios of transition to “unsustainability”, and ways of responding to them. “Such a study may help to overcome a lack of interest in this question, something which was criticized by several speakers,” say the authors of the statement.

“We should engage neither in Cassandra-like announcements of future catastrophes, nor in any irresponsible optimism,” the authors write. “Today, in the face of the global complexity of our contemporary context, the human being, more than ever before, is called upon to find that right kind of rationality which will achieve survival and sustainability through the application of new and well-deployed practical criteria.”

Full text: <http://helix.nature.com/wcs/m16s.html>

India's science funds 'safe' despite fall of government

[NEW DELHI] India's senior science administrators say that the fall of the government last week is unlikely to have a significant impact on science and technology activities, provided that the budget is passed without cuts by the new government.

The 1999–2000 budget was presented by the coalition government led by the Bharatiya Janata Party (BJP) on 28 February (see *Nature* 398, 4; 1999). The budget must be approved by parliament within 75 days, but the government was defeated in parliament in a confidence motion last week. Opposition parties led by the Congress party are likely to pass the budget without making major changes.

“The BJP had given us an excellent budget for science this year and we never had it so good,” says Ragunath Mashelkar, chief of the

Council of Scientific and Industrial Research. “I hope the budget will be passed and that there will not be any cut.”

Some observers say, however, that, although the budget for science is unlikely to be revised, the allocation for the nuclear weapons programme — a top priority of the BJP government — may be cut by the Congress party if it forms the government.

Valangiman Ramamurthi, secretary of the ministry of science and technology, says the change in government is of no significance as far as India's participation in the World Conference on Science is concerned. India's position at the conference has been decided, although “who will lead the delegation is something we have to wait for until a new government is formed”.

Full text: <http://helix.nature.com/wcs/b31.html>

Canadian plea for freedom of information

[OTTAWA] A group of graduate students at Carleton University in Ottawa, Canada, has produced a set of 21 ‘principles for science’, intended as a contribution to the discussion leading up to the World Conference on Science, and described as “a declaration of interdependence for the 21st century”.

The principles endorse the concept of scientific freedom — including the freedom to publish research results and attend conferences at home and abroad — while acknowledging that such freedom “is neither absolute nor unfettered”. They also argue for a better, long-term commitment to basic science in all countries.

They state that, contrary to trends

towards commercialization, scientific knowledge should be disseminated freely and openly. “Developing countries, in particular, require affordable and ready access to research results if they are to prosper in the next millennium.”

The students are participants in a graduate seminar on science, technology and international affairs, led by Gerald Graham of the Norman Paterson School of International Affairs.

One of the more controversial suggested principles is that “the public has a right to be involved in every stage of the scientific process, from the conception of a project to its ultimate completion”.

Full text: <http://helix.nature.com/wcs/a25.html>

Intellectual rights ‘must serve the needs of developing countries’

[PORT OF SPAIN] The international community needs to ask itself whether current forms of intellectual property protection are appropriate for the needs of the developing world, says Kamla Persad-Bissessor, minister of legal affairs for Trinidad and Tobago.

According to Persad-Bissessor, in the rush to appear developed, developing nations may have emphasized the acquisition of consumer durables rather than building the infrastructure that could have sustained further development. In

doing so, she says, they took on the products of the research of developed nations without doing the research themselves, often feeling free “to borrow, copy and plagiarize the knowledge of developed nations”.

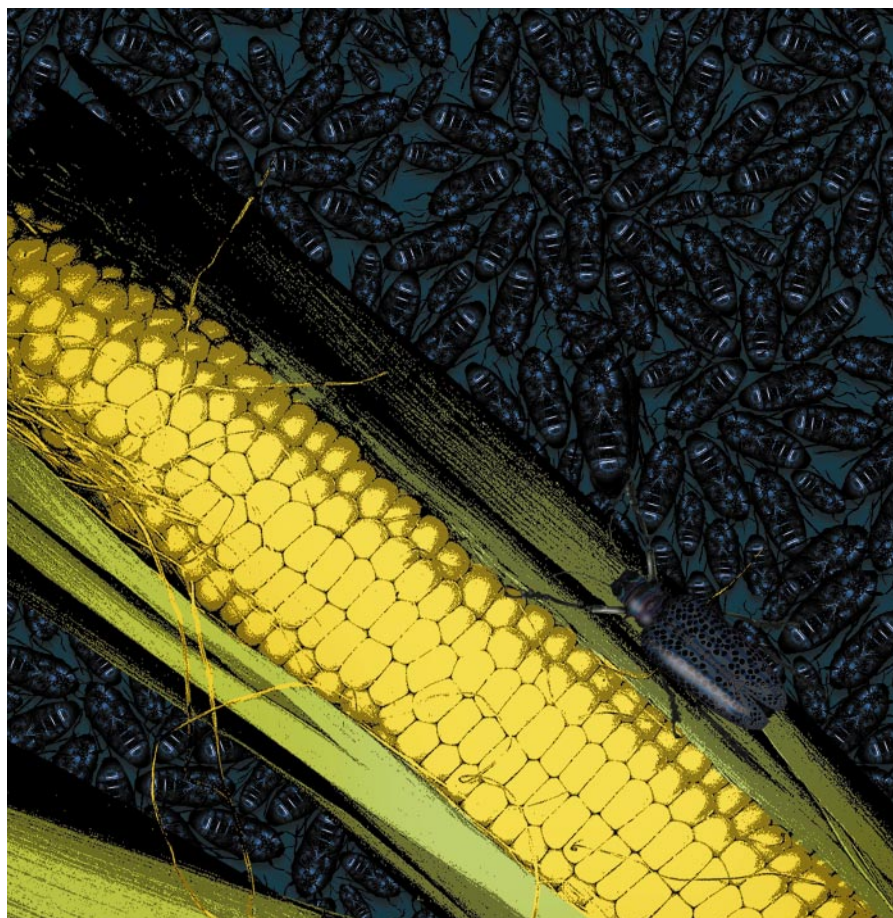
In developed countries, she points out, the growth of intellectual property protection occurred alongside technological and other developments. But, in developing countries, intellectual property protection has in a sense been imposed by the need of the first world to protect its interests.

“This provides an interesting challenge

to the legislators of developing nations, to ensure that model legislation provided obligingly by the world Intellectual Property Organization is appropriately tailored to the needs of the local creators and inventors,” said Persad-Bissessor.

Speaking at a workshop organized by the Caribbean Academy of Sciences, she said that certain provisions might have to be modified, or new areas of protection developed within the legislation, to meet the need for development.

Full text: <http://helix.nature.com/wcs/c11.html>



Long-term effect of GM crops serves up food for thought

The media has inflamed public fears about the risks of genetically modified crops for human health and biodiversity. But many responsible scientists agree on the need for more research to identify potential long-term problems.

The declared position of the world's major regulatory and scientific agencies is that, in principle, genetically modified (GM) crops pose no greater threat to human health than those produced by traditional breeding.

"I don't see any problems at all for genetically modified plants in terms of human health. Researchers are being asked to prove negatives," says Robert McKinney, director of the division of safety at the US National

Institutes of Health in Bethesda, Maryland.

But this view is being seen increasingly as an oversimplification by critics of gene technology — and by some of its supporters. They argue that it could stifle serious consideration of the 'remote but real risks' potentially associated with the acceleration and broadening of plant breeding brought about by the development of genetically modified plants.

Most scientists believe that such risks are largely hypothetical, and that current safeguards are adequate. Even among ardent supporters of GM foods, however, calls are being increasingly heard for more research on health risks, and for the introduction of monitoring systems that would allow the

early detection of any long-term problems.

For example, although the question of whether to label GM foods is usually considered to be an issue of consumer choice rather than public health, several scientists say there is also a strong argument for labelling to facilitate epidemiological studies to detect any increases in allergies or diseases that might be linked to GM foods.

The need for careful monitoring is urgent, given that the introduction of thousands of GM foods on a global scale appears imminent, says Suzanne Wuerthele, a risk assessor at the US Environmental Protection Agency, speaking in a personal capacity.

This view is supported by Ben Mifflin, former director of the Institute of Arable Crops at Rothamsted, near London, who is a proponent of the potential benefits of genetic modification of crops. He argues that, under current monitoring conditions, any unanticipated health impact of such foods would need to be a "monumental disaster" to be detectable.

Mifflin points out that a general increase in gastrointestinal disorders, for example, would be difficult to attribute to a particular food, given the diverse possible origins of such symptoms. "So, yes, there is an advantage for going slowly."

Some researchers have proposed specific monitoring strategies. Hans-Jörg Buhk, director of the Robert Koch Institute in Berlin, has called for the creation of a 'gene register' to track which genes and constructs have been introduced into the crop gene pool. This precautionary measure, he argues, would improve the ability of researchers to predict interactions between genetic modifications.

Whereas the genetic traits currently being introduced tend to be relatively simple, monogenic ones, the next generation of GM crops may require a rethink, says Buhk. These could include 'functional foods', such as plants with increased vitamin levels, or pharmaceutical-producing plants. He predicts that, before the full-scale introduction of such plants, additional safety trials may be needed, perhaps analogous to the clinical trials used to assess the safety of drugs.

Scientists "do not know everything in advance," says Buhk, adding that unexpected events cannot be ruled out. He points to a tomato developed by the British company Zeneca that was designed to have an increased shelf life, using antisense technology directed against the polygalacturonase gene that causes ripening. According to Buhk, the best performing line was in fact caused by an unpredicted 'sense' event (gene activation). "This was a rare event, either a contamination or a chance turnaround [in the genome]," he says.

In a similar way, virus-resistant plants

This Briefing has been written by Declan Butler and Tony Reichhardt with additional reporting by Alison Abbott, David Dickson and Asako Saegusa.

On other pages | [Research goes back to basics: p652](#) | [Biodiversity in the balance: p654](#) | [Liability warning: p656](#)

have been genetically engineered using an approach where expression of a viral coat protein confers a virus-resistant phenotype on the plant. This strategy has been demonstrated against many viruses. But it has been occasionally observed that expression of the viral coat protein is unnecessary, and that translation of parts of the gene alone can confer the desired protection. "This is gene silencing," says Buhk. "There is interaction going on at the RNA level that we do not understand."

The likelihood and impact of such unexpected occurrences seem to be the major area of uncertainty among scientists working on the potential health impacts of GM foods. At the same time, many are highly critical of recent health scares, claiming that the arguments for and against such foods have been obscured and exaggerated, if not distorted.

The general view of such researchers, for example, is that, although the safety of antibiotic marker genes has received prominent media attention, it is ultimately a secondary issue as far as potential impact on human health is concerned. Whatever the perceived risks of such markers, there is a broad consensus that their use is unnecessary, and that innocuous alternatives are available.

Another public concern in genetic engineering in general is the safety of the vector used. But many scientists believe that the genetic constructs used to engineer plants — such as the *Agrobacterium tumefaciens* vector and the cauliflower mosaic virus (CMV) promoter — are inherently safer than, for example, the attenuated viral vectors used in gene therapy.

"You get massive doses of CMV each time you eat broccoli," says Marc Van Montagu, professor of genetics at the University of Ghent in Belgium.

What is clear, however, is that crop plants contain many compounds that have the potential to damage human health, and that even conventional plant breeding has on more than one occasion come close to producing plant varieties that could pose serious health risks.

"There are several well documented examples where breeding has inadvertently introduced higher toxin levels," says Mifflin, pointing to cases of high levels of psoralen in celery, and of solanine in potato. "There are hidden risks in any form of plant breeding."

Such toxic varieties never came to market, however, as they were detected by the quality control procedures of the seed companies themselves. Mifflin and others believe that, if anything, the generally stricter approval procedures required for crops produced by genetic modification reduce the likelihood of such mishaps.

But this does not reduce the need for vigilance. Nor of the importance of research to better understand the impact of introduced genetic constructs on plant metabolism, and

Basic research holds key to weighing risks

The apparent low level of studies into the long-term risks of plant biotechnology seems to result partly from the relatively small sums devoted to risk assessment research in comparison to overall budgets for biotechnology.

Neither the industry nor most governments set aside specific sums for biosafety research, which is regarded in countries such as Britain and France as part of biotechnology research. The US Department of Agriculture, in contrast, puts aside one per cent of its biotechnology research budget for biosafety projects — around \$1.5 million a year.

Germany, with its history of hostility to genetic engineering, spends six per cent of its biotechnology research budget — DM10 million (\$5.5 million) — on biosafety studies. This sum will increase substantially this year, following a pre-election promise

by the new Social Democrat–Green coalition government.

Uncertainty remains in Germany over how the extra funds should be spent. But likely priorities are monitoring the effects on gene flow and local ecology of crops grown on a commercial scale — in common with trends occurring in other countries (see page 655). Extra funds are also expected to be committed to such research by the British government, in response to recent public anxiety over the safety of GM food.

Many scientists, however, say they are not overly concerned about the relatively low levels of funding. John Beringer, a microbial geneticist at the University of Bristol, and long-serving chairman of the British government's Advisory Committee on Releases to the Environment, says his committee has never been refused money for research that it recognized as important.



Growing problems: fundamental research will help make genetic engineering more predictable.

to improve existing techniques for evaluating toxicity and allergenicity.

Underpinning health risk assessment throughout the world is the concept of 'substantial equivalence', a term coined by the Organization for Economic Cooperation and Development in 1993 under which GM foods are compared with analogous conventional foods in terms of toxicity, nutritional qualities and other characteristics.

In line with this approach, applications to the UK Advisory Committee on Novel Foods and Processes require companies to show that their products are 'substantially equivalent' with respect to known toxins and allergens. A company submitting a modified potato would need to show that the genetic modification had not inadvertently increased alkaloid levels, for instance.

The first issue addressed by approval procedures is the safety of the introduced genes — for example their purity — and the proteins they express. "If you introduce a gene

for a poison into strawberries, you shouldn't be surprised if the plant is toxic," quips Axel Kahn, former head of the French committee that approves GMOs.

Where proteins might be degraded by plant enzymes, the toxicity of any breakdown products is also considered in this procedure. Introduced genes coding for enzymes can in principle also modify the metabolism of plants, leading, for example, to an increase in toxins. Mutagenesis by genes randomly inserted into the genome also raises the risk that 'silent' or lowly expressed genes might be switched on or upregulated with damaging results.

The need to assess the potential risks of such unpredictable occurrences cannot be ignored. Plans to market *Canola* and *Vicia* (bean) plants, genetically modified to boost their low cysteine and methionine content using the methionine-rich 2S storage albumin of Brazil nuts, were abandoned when it was discovered that the protein was highly

Beringer, in common with colleagues in other countries, also believes that basic research into how exactly genes are controlled, rather than research intended to answer specific problems, will be the key to addressing biosafety issues in future. Directed research programmes are often not best suited to assessing the potential risks of plant biotechnology, says Beringer. "We don't have very precise questions, so it would be difficult to formulate too exact a research strategy."

Phil Dale, research group leader at the John Innes Centre in Norwich, England, worries that politicians and government officials may feel under pressure to embark on programmes designed primarily to allay public fears. In his opinion, "what we really need to control safety is a much better understanding of what influences gene action generally".

One example of the type of unexpected phenomenon requiring greater understanding is that the orientation of the long-shelf-life gene introduced into Zeneca's genetically modified tomato turned out to be the opposite of what was intended (see page 651).

Lothar Willmitzer, director of the Max Planck Institute for Molecular Plant Physiology near Potsdam in Germany, agrees with Dale that the ecological risks of GM crops cannot be understood in isolation from more basic questions in plant biotechnology. "It is nearly impossible to design meaningful programmes."

At present, risk assessment research in agricultural biotechnology is primarily driven by regulatory requirements, based on the fact that companies wishing to

undertake field trials need to provide detailed information on inserted genes, their expression and stability.

However most countries also run research programmes of broader relevance, relating to more general problems identified by ministries' scientific advisory committees. These include literature searches, which bring together existing information from scattered sources. Research agencies also fund individual projects in biosafety. Research policy makers such as Beringer believe that this approach works well.

Some environmental pressure groups also have no difficulty with such a strategy, although most oppose farm-scale experimental trials, mainly because they are seen as representing a step towards commercialization, which all environmentalist groups oppose, regardless of research outcomes.

Green groups are also concerned at the relatively large numbers of government scientific advisers in countries such as Britain with links to the biotechnology industry. Biosafety is being driven by commercial interests, says Peter Riley of Friends of the Earth UK. He adds that "we need to get down to basics and hand money freely to universities to do fundamental research and make the science of genetic engineering predictable".

Industry appears to be in agreement with this sentiment. Nigel Poole, regulatory affairs manager at Zeneca, says companies are willing to help pay for both basic and directed research, providing it is of the highest quality. "A strong knowledge base is essential," he emphasizes.

allergenic (see *The New England Journal of Medicine* 334, 688–692; 1996).

Regulatory authorities are confident that the allergenicity of untested proteins can usually be reliably predicted by structural analysis and testing. Most allergenic proteins tend to have a molecular weight between 10 and 70 kilodaltons, often share characteristic sequence stretches, and resist degradation by heat, as well as acid and pepsinase conditions mimicking those found in the stomach. Some scientists, however, believe that regulatory agencies overestimate their ability to predict allergenicity.

More difficult is the assessment of the potential presence of unknown toxins. Animal feeding experiments are plagued with problems, such as high levels of variation in results between experimental groups. It is also difficult to feed relevant doses of GM foods to laboratory animals for sufficient periods to allow reliable extrapolation of the results to humans.

Another problem, as several scientists point out, is that such research is unattractive to researchers, as it tends to yield negative results that are difficult to publish and to account for to funding agencies.

A potential alternative to feeding experiments is to subject plants to a battery of chromatographic and spectroscopic analyses, to create a molecular 'fingerprint'. Any difference between the fingerprint of a modified crop and an existing one would indicate a need for more extensive safety testing.

However efficient the techniques used to pinpoint potential problems, critics such as Wuerthele, and many consumer groups, remain sceptical of the reliability of safety assessment procedures, and the judgement of expert committees — especially following the failure of experts to prevent the bovine spongiform encephalopathy crisis.

Foreign proteins that have never been in the human food chain will soon be consumed in large amounts, she points out. "It

took us 60 years to realize that DDT might have oestrogenic activities and affect humans, but we are now being asked to believe that everything is OK with GM foods because we haven't seen any dead bodies yet."

Many researchers point out that such unknowns are common in classical plant breeding, however. "I honestly don't see what you could do to a plant that would be so different to the traditional approach," says Mike Gale, director of the John Innes Centre in Norwich, England. He argues that, although transgenes are incorporated in a random fashion into the plant genome, the introduced gene is well characterized.

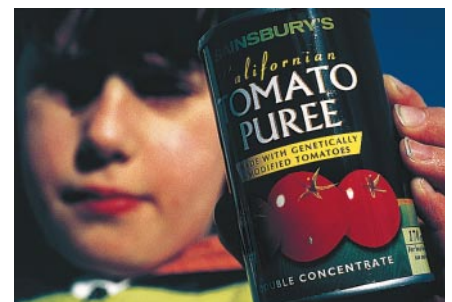
In contrast, in conventional breeding, Gale points out, in addition to introducing a desired trait into a crop from a wild relative, breeders have no idea what other changes they may have introduced through the integration of large chunks of the donor genome. "Surely putting in one gene is better than shuffling around tens of thousands at random," he says.

Similarly, genetic engineering may actually help to improve food safety by removing toxic compounds naturally present in many plants. Cassava, a major staple food, contains cyanogenic glycosides, for example, that must be eliminated by fermentation before the plant can be eaten. Kidney beans require soaking to remove toxic lectins — poor food preparation still results in deaths.

A major goal of plant breeding throughout the ages, points out Mifflin, has been to reduce or eliminate the many toxic compounds present in wild plants that have evolved, for example, as defence mechanisms against predators.

But, whatever the validity of such arguments, there is growing awareness — even among major agrofood companies such as Monsanto and Novartis — that consumer concerns about health cannot simply be wished away if the fruits of biotechnology are to be realized.

"It is clear [at least] in Europe that the consumer backlash means that more investment in research [and monitoring] is needed to clarify the scientific uncertainties," says Gerry Moy, head of food safety at the World Health Organization. "Society needs to begin a broad discussion on the potential risks and benefits of GMOs."



Backlash: consumer concerns forced action on labelling and could now influence research goals.

NICK COBBING/STILL PICTURES

Assessing the threat to biodiversity on the farm

The unintended consequences of genetically modified agriculture for the preservation of biodiversity have long been the focus of international attention, perhaps raising even more controversy than the potential impact on human health.

Talks on a new global treaty to minimize the possible adverse environmental impacts of GMOs broke down earlier this year, for example, partly because governments, environmentalists and industry disagree strongly over the ecological risks of GM crops (see *Nature* 398, 6; 1999).

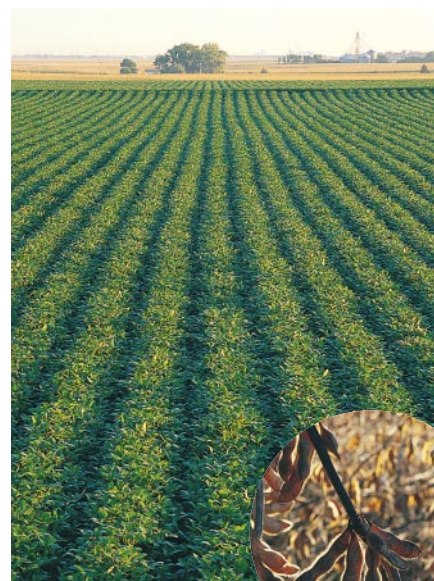
This disagreement is fuelled by the patchy state of research on the issue. Although there is plenty of evidence that modern farming methods have reduced biodiversity in many countries, a report by the British government's advisers on GM releases expresses a widely held view when it says that researchers do not yet know whether the planting of genetically modified crops will make things better or worse (see <http://www.environment.detr.gov.uk/acre/wildlife/index.htm>).

These issues have perhaps been of more concern to small countries such as Britain, where most of the countryside is used for agriculture, than to larger ones with wide

expanses of land that is not farmed. Growing protests from environmentalist and conservation lobbies have persuaded the UK government and the scientific community to embark on a comprehensive programme of research on the impacts of GMOs on wildlife and the countryside.

In contrast, there has been less consciousness of these research needs in the United States. This is partly because of a perception that such research is not required, as the distances between farmland and the wider countryside are much greater. But it may also be because the US biotechnology industry, the world's largest, has had greater success in conveying to the public the message that GMOs are environmentally friendlier than conventional agrisystems.

Most GM crops already in use have been modified to confer tolerance to herbicides or to insects (although many other varieties of GM plants have also been commercialized). Herbicide-tolerant plants, in general, are modified to resist the commonly used herbicides glyphosate and glufosinate, which can therefore be sprayed on crops without damaging them. Insect-resistant plants are modified to produce toxins made by the soil bac-



Field trials — such as these of soyabean in Nebraska — are essential.

terium *Bacillus thuringiensis* (Bt) that kill specific target pests because of the interaction between proteins produced by the bacterium and the pest.

Whether insect-tolerant plants can in practice harm non-target insects — and birds and mammals — is high on the list of questions for multi-year farm-scale experiments with GM crops that are being planned

A pragmatist comes to the defence of the British countryside

No-one would lightly accuse Sir Robert May, the British government's chief scientific adviser, of being a romantic. The Australian-born theoretical-physicist-turned-population-biologist has established a powerful reputation for applying mathematical modelling to problems ranging from the preservation of biodiversity to the spread of AIDS.

His pragmatism has been well to the fore during the recent British controversy over the potential health effects of GM foods. A willingness vehemently to criticize the unscientific nature of many of the claims being made turned him into a key spokesman for the government. Equally passionate was his dismissal of the popular newspapers whose reports were fanning the controversy over so-called 'Frankenstein foods' as "straight entertainment".

Where May does have concerns is about the long-term implications of GM-based agriculture on biodiversity. He points out that the history of agricultural change is in the direction of growing crops "that no one eats but us". This has obvious consequences for the animals that also depend on the fields we use. May quotes, for example, recent surveys of the decline of many bird

populations, and says he is convinced by evidence for corresponding effects on invertebrate and plant diversity. "The thrust of GM crops is to accelerate this trend."

May admits that there remain scientific uncertainties about the health and environmental effects of GM food and crops. But he has little time for the claims by Scottish researcher Arpad Pusztai to have detected a depressed immune response from eating potatoes genetically engineered to produce the toxin lectin.

"That is not scientific uncertainty; as long as it remains unpublished, it is outside the canon of science," says May.

May is confident that, if there had turned out to be such a danger with existing GM foods, Britain's regulatory authorities would have picked it up. But he also points out that the experience with bovine spongiform encephalopathy has brought home the need to expect the unexpected. "We must test," he wrote in a paper for the Prime Minister Tony Blair. "No-one was looking for untoward effects in cattle. In the case of GM food, we are testing for unexpected and unwanted effects on health and the environment."

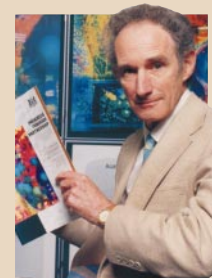
Some widely quoted risks are, he says, relatively low, even if undesirable. One is the

spread of 'superweeds' resulting from the interbreeding of herbicide-resistant crops with wild relatives. Although admitting that this could happen, he argues that the weeds would remain vulnerable to other herbicides.

Of slightly greater concern, he says, is possible cross-pollination with other crops. "We need to know rather more about this than we do at present."

His real worry, however, is about the impact of GM crops on biodiversity. "We need wider mechanisms to reconcile farming with preservation of the countryside," he says. "There is a much larger role for trying to understand how subsidies and other instruments interact with environmental protection."

But May is adamant that the remaining uncertainties provide a basis for field trials "on a scale sufficient to answer the questions we face" — not for a moratorium on trials.



May: concerns for the countryside.

by the British government. Another question to receive close scrutiny will be the extent to which modified genes can be transferred to other plants, and what effect this might have on, for example, organic produce. A third question is whether herbicide tolerance can spread to nearby plants, whether weeds or other crops.

The limited evidence available so far has left researchers apparently divided on the risks to non-target insects from *Bt* crops. William Hutchison and colleagues from the department of entomology at the University of Minnesota, for example, told a meeting of the Entomological Society of America last month that they found no difference in the numbers of 'beneficial' insects when they sampled fields of *Bt* sweetcorn and non-*Bt* corn in Minnesota (see <http://www.ent.ias-tate.edu/entoc/ncb99/prog/abs/d51.html>).

In contrast, Nicholas Birch, a research entomologist at the Scottish Crop Research Institute in Dundee, has demonstrated in lab studies that an anti-aphid toxin expressed by an experimental GM potato reduces the fertility and shortens the lives of ladybirds that eat the target aphids. Critics, though, point out that the toxin in question, snowdrop lectin, is unlikely to be approved for a GM crop given previous evidence of its toxicity.

Perhaps a more realistic pointer to potential dangers has come from Angelika Hilbeck, of the Swiss Federal Research Station for Agroecology and Agriculture in Zurich, who has found that lacewings, another beneficial insect, have higher death rates when fed the larvae of target insects that have eaten *Bt* corn compared with larvae fed on ordinary corn.

But Hilbeck's studies were conducted in the laboratory (see *Environmental Entomology* 27, 480–487; 1998). Under farm conditions, the results may be different, as the target insect — the European cornborer — lives inside corn stalks, where under normal conditions it is largely protected from lacewings.

Research has also been under way for some time to assess the impact on nearby flora of herbicide-tolerant GM crops. Researchers from Denmark, France and the United States have already suggested that the results of trial experiments indicate that herbicide-tolerant genes can in principle 'escape' from GM plants to nearby weedy relatives through pollen transfer.

Similarly Anne-Marie Chevre, a plant researcher at the Institut National de la Recherche Agronomique in Le Rheu, France, has found that oilseed rape genetically modified to withstand herbicide easily produced fertile offspring when crossed with a common weed, the wild radish — although she also found that the herbicide-tolerance genes became more diluted with each generation of hybrids.

Scientists working for environmentalist groups are among those who claim that such

Japan defends its drive for self-sufficiency

Despite widespread public concern over genetically modified food, Japan's scientists — in concert with those in other countries — have only recently begun to address questions on the potential long-term risks to human health and the environment from GM crops.

On 1 April, the Ministry of Agriculture, Forestry and Fisheries (MAFF) embarked on the government's first project to examine the risks from GM crops. As elsewhere, the project will focus on the long-term impacts of herbicide- and insect-tolerant crops on ecology and on agricultural practices.

A substantial proportion of the Japanese public, like their counterparts in Europe, are uneasy about GM foods. Some invoke ethical concerns about the manipulation of genes. Another reason is a relative lack of public understanding of genetic modification techniques. A third reason is government reluctance to label GM foods.

But other factors may be at work. According to Naoto Shibuya, a researcher in bioengineering at the National Institute of Agro-Environmental Sciences in Tsukuba, public unease can partly be attributed to an absence of effective public communication of the risks. Shibuya says scientists need to communicate in a way that "indicates what is understood and what is not".

A low level of public confidence in GM foods is not good news for the government, which is relying on GM agriculture to make Japan self-sufficient in food. Unsurprisingly, finding ways of allaying public concerns is a key aim of the MAFF research project.

Japan's GM regulations are modelled on a framework set out by the Organization for Economic Cooperation and Development (OECD), with a strong emphasis on the concept of 'substantial equivalence' (see page 652). But critics question 'substantial equivalence' as a basis for deciding whether a product is safe, and argue that there is no



Farm tests: Japan has tough rules on field trials.

substitute for long-term risk assessments. "The regulators have overlooked the potential residual toxicity after several growing seasons, and the consequences on genetic diversity," says Setsuko

Yasuda, director-general of Japan's Consumers' Association.

But representatives of the Ministry of Health and Welfare, as well as MAFF, which are both involved in the approval of GM products, claim that the chances of GM crops posing health and environmental risks are negligible, and that such risks would be detected during safety tests.

In some ways, Japan's regulatory system for GM crops is tougher than in other industrialized countries. Once a potentially useful crop plant has been developed, small scale, isolated field trials are carried out, followed by cultivation over at least one generation in a farm-scale environment. Farm-scale trials are not a regulatory requirement in many OECD countries.

Although the government is keen to press ahead with the development of GM crops, the private sector is more cautious than in other industrialized countries. Companies such as Japan Tobacco, Kirin Beer and Suntory are carrying out farm-scale trials — including virus-resistant rice and petunia — but there are no immediate plans to commercialize any of these products.

According to government sources, this is because few locally-based companies are keen to be the first to commercialize GM crops because of a fear that this could create a negative image of the company, and perhaps trigger a boycott of its products.

gene transfer could encourage the proliferation of 'superweeds', which might turn out to be highly invasive.

The consensus from a recent gathering of scientists, regulators and research managers in Bethesda, Maryland, convened to consider the ecological impacts of GM crops, was that there is little risk of enhanced weediness from the handful of transgenic plants on the market. Their genetically enhanced traits would not confer any competitive advantage over other plants, and would eventually die out, it was concluded.

But overall the jury is still out. For example, the scientists attending the Bethesda meeting agreed that, when many different GM plants exchange genes, a kind of 'gene stacking' of multiple desirable traits could

theoretically produce a highly competitive weed. And some, such as Allison Snow, an ecologist at Ohio State University, point out that this is not likely to be known until many GM crops are in wide use.

Nor do researchers yet know whether a fitness-improving gene — such as one that confers resistance to pests, herbicides or drought — will necessarily make a weed or a GM crop more invasive. In the case of herbicide resistance, unless the weed is sprayed with herbicide, there should be no selection pressure favouring the survival of resistant plants, and the trait should die out in time.

Representatives of the biotechnology industry are among those who believe strongly that the benefits to agriculture and the environment from GM crops tend to be

understated in debates on the impacts of such crops.

One such benefit is that insect-tolerant crops need smaller quantities of conventional pesticides, whose ability to harm the environment is well documented. Indeed, a report by the privately financed National Center for Food and Agricultural Policy in Washington D.C. points out that pesticide applications in southern US states have dropped significantly in recent years, coinciding with the spread of *Bt* crops (see <http://www.ncfap.org/biotech/sld014.htm>).

Despite such a benefit, however, the use of *Bt* as a spray, as well as what has been described as an exponential increase in 'Bt plants', has raised concern that target insects could eventually become immune to the toxin, a scenario that would harm both agriculture and the environment.

One exponent of this view is Bruce Tabashnik, an entomologist at the University of Arizona. Tabashnik admits that there are no well documented cases of pests becoming resistant to *Bt* crops. But, in common with many farmers, he believes that resistance is inevitable. So far, Tabashnik and others have shown that one species, the diamondback



moth, has shown widespread resistance to *Bt* spray. And Fred Gould, an entomologist at North Carolina State University, has found higher than expected frequencies of alleles conferring resistance to *Bt* in field populations of the tobacco budworm (*Proc. Natl Acad. Sci. USA* **94**, 3519–3523; 1997).

Gould, in common with other researchers and farmers, says that methods need to be found to deal with the risk of pests becoming immune to *Bt*. One method being proposed is for farmers to deliberately set aside more field space for growing non-*Bt*-cotton. The idea behind such 'refuges' would

be to allow budworms to breed with *Bt*-exposed pests, so diluting *Bt*-resistance genes in future generations of budworms.

There appears to be a consensus among corn growers, researchers and biotechnology companies that at least 20 per cent of the growing area should be set aside for non-*Bt* corn in this way — farmers currently set aside just four per cent. The loss of *Bt* as an effective pesticide would not necessarily pose a new ecological hazard; resistance to pesticides predates GM crops. But it could make biotechnology's victory in increasing food production a short-lived one.

Scare tactics: talk of 'Frankenstein foods' may be exaggerated and misleading. But the biotech industry has still been forced to respond to many of the concerns of consumer groups and environmentalists.

Industry critic warns that damages claims 'could run into millions'

"It's an 'emperor has no clothes' situation," says Jeremy Rifkin, one of the biotechnology industry's most vocal critics. "You cannot have governments telling us that the technology is safe when there is no science to judge it by."

The absence of what he calls a 'predictive ecology' will, he suggests, have a direct impact on the industry: insurance companies will be reluctant to issue protection against claims for environmental damage if there is no way of quantifying what this damage might be.

"You're going to see lots of litigation when genes start flowing to organic crops, or to wild relatives on neighbouring lands," he predicts. "The gene flow is going to be on a scale that people have not understood. Liability is going to be the Achilles' heel of the biotechnology industry."

Rifkin has headed a small but influential Washington-based pressure group, the Foundation for Economic Trends, since the late 1970s. His apocalyptic scenarios have won him few friends — and many enemies — in the biotech community.

But the themes he emphasizes — in particular, that genetic engineering is somehow 'unnatural', and almost by definition potentially dangerous — have hit a responsive chord among the public. And it is the implications of that for indicating the direction of consumer demands, such as growing insurance claims, that have given him a ready audience in US boardrooms.

Rifkin has a clear sense of where he sees future problem areas. With herbicide resistance in plants, for example, he identifies the issue of such resistance spreading out of control. "The industry argues that inserting a herbicide-resistant gene will mean more sustainable agriculture, but it could be the opposite. If you put in a herbicide-tolerant plant, and then increase the use of that herbicide, how long will it take for resistant strains of weeds to appear?"

"You see it more urgently in the case of pest resistance. Here you could end up with every cell of every plant producing a toxin. Because it is only a single gene, the 'magic bullet' runs out very quickly. It is faster and easier for an insect to build up resistance and, the more widely these plants are used, the greater the problem is likely to be."

Even if no overt damage occurs, Rifkin argues that gene flow into neighbouring crops is likely to become a source of conflict. "Foreign genes are a 'smoking gun'. They are going to flow all over the place, and they will always be identifiable."

"Claims for damages could come from gardeners or organic farmers who find they are unable to sell their crops. All that has to happen is for a gene to turn up that you did not want. The overall claims for damages could make the recent litigation associated with smoking pale in comparison."

One option for the industry, he suggests, is to turn to the government for financial protection. "But I don't know anyone in the



Rifkin: liability will be biotech's Achilles' heel.

world who will allow it to happen for biotechnology. There is a potential vulnerability here that is so dramatic and so unaddressed that it cries out for attention."

Hence the need for a predictive ecology

— to provide financial security, if nothing else. "At present, the insurance industry is not likely to want to touch this type of thing; you have to have predictability."

The same issue applies to potential health impacts. "We just do not know, if you take a gene from an unrelated species that codes for a protein that has never been part of our diet, what the allergenic impact is likely to be."

"It does not take much imagination to suggest that not all the genes that code for proteins are going to be safe. And, given the scale on which these foreign genes are being introduced into foods, I predict that there could be quite a bit of illness — another issue that is going to force the liability question."

Scare tactics, perhaps. But, given the extent to which the regulatory agenda is set by the reaction of politicians to public sentiment, Rifkin insists that these issues need to be taken seriously. He stresses the need for "serious research" into long-term, low-level consequences: "This is essential if the industry is going to survive." □

When extra authors get in on the act

Sir — You report that more than two-thirds of Europe's young scientists say they are not given full credit for their research achievements¹. There are similar concerns in the United States. A survey of 191 US postdoctoral physicists finds that senior scientists are frequently listed as authors of papers even though they have had little or no participation in the work². How does this come about?

The results indicate that there is no standard process for the distribution of authorship among co-workers. The American Physical Society's ethical guidance says: "Authorship should be limited to those who have made a significant contribution to the concept, design, execution and interpretation of the research study." Some 76 per cent of the survey's respondents had not seen the ethical statement. When shown it, half of the postdocs said that, according to their interpretation of the statement, obtaining grants and funding would be a qualification for authorship. (The statement does not specify that authorship contributions should be intellectual, nor original.)

Seventy-five per cent of postdocs had

never discussed authorship criteria with their supervisors, and 70 per cent said the criteria for designating others as authors were not "clearly agreed upon".

Postdocs perceive there to be a substantial amount of inappropriate authorship. Supervisors were authors on 92 per cent of postdoc papers. Guided by the ethical statement, the postdocs responded that in 14 per cent of those papers the supervisor should not have been listed as an author. In 33 per cent of papers with authors in addition to the supervisor or postdoc, one or more of those authors should not have been listed. Respondents reported that in one per cent of papers they were themselves inappropriate authors.

The reasons reported for inappropriate authorship fell into four groups:

(1) Concern about the relationship between postdoc and supervisor. Postdocs need letters of recommendation from supervisors and want to keep in their good graces. Relationships with other scientists in the field are perceived to be enhanced by giving them authorship. Sometimes the authors hope to gain prestige or expedite publication by adding a well-known name.

(2) Minor contributions to the work, more appropriate for acknowledgement than authorship.

(3) Previous work in the field, or expected contributions that did not materialize.

(4) Crediting staff who are socially close, for example part of the same research group.

The scientific community should adopt formal procedures for assigning authorship appropriately and penalizing 'honorary' authorship. This would avoid the problem that it can be difficult to discuss credit within a research group. Authorship of papers could be modelled on US patent procedures in which an independent party inquires into the work and assigns authorship. Or a section could be added at the end of each paper to explain what each author contributed.

The impetus to introduce these changes lies with junior scientists, the most likely victims of having their intellectual property 'diluted' by inappropriate authorship.

Eugen Tarnow

Avabiz.com, 3725 Blackstone Avenue, Ste. 4E,
Riverdale, New York 10463-1444, USA

1. Schiermeier, Q. *Nature* **397**, 640–641 (1999).

2. Tarnow, E. *Science and Engineering Ethics* **5**, 73–88 (1999).

Miscalculation

Sir — Your headline "Germany enters the supercomputer age" is some 20 years out of date (*Nature* **397**, 643; 1999). It is even more ironic as the Rechenzentrum Garching of the Max Planck Society was the first centre in the world to offer civilian scientists open access to supercomputers, inspiring Larry Smarr and his colleagues to found the US national supercomputer centres.

Mordecai-Mark MacLew

Max Planck Institute for Astronomy,
Königstuhl 17, D-69117 Heidelberg, Germany

Hand over your clones or lose your reputation

Sir — Noel Harris is right to protest the refusal of many researchers to respond to requests for clones, especially having signed away their right not to do so by publishing in journals such as *Nature* (*Nature* **398**, 102; 1999). But, rather than 'outing' such workers, it is much more effective simply to clone the gene yourself and then do better work, faster, than those who tried to slow you down (driven by the incentive of revenge?).

Given the speed with which known sequences can now be isolated, refusal to

part with clones is just plain silly — the cost in terms of ill-feeling, with or without 'outing', far outweighs the few weeks' benefit gained over competitors. The 'refusers' are already losing out.

Mark Tester

Plant Sciences Department, University of Cambridge,
Downing Street, Cambridge CB2 3EA, UK

Did Parisians catch HIV from 'monkey glands'?

Sir — It was good to read Gao *et al.*'s report¹ which now leaves little doubt that chimpanzees (*Pan troglodytes troglodytes*) from central Africa are the origin of the human HIV-1 pandemic. Two additional points seem worthy of comment.

Chimpanzees live in multi-male communities with a promiscuous mating system, quite unlike humans or the other great apes. This makes them particularly susceptible to the spread of sexually transmitted infections, so they might have evolved some novel defence mechanisms which have allowed them to come to terms with HIV-1. If only we knew how.

One of the consequences of the chimpanzees' promiscuous mating system is that they have developed extremely large testes in response to gamete selection².

Unfortunately, this made chimpanzees attractive subjects for Serge Voronoff, director of experimental surgery at the Collège de France in Paris. In the 1920s, Voronoff was the leading exponent of testicular transplantation for the rejuvenation of ageing men^{3,4}. Lacking an adequate supply of fresh human testes courtesy of the guillotine, he resorted to chimpanzee testes, which had the added advantage that each was so large that it could provide scrotal transplants for a number of men. Since his chimpanzee donors could well have come from Francophone central Africa, it is quite possible that some of his elderly male patients, and maybe even their partners, might have acquired an HIV-1 infection and developed AIDS.

Perhaps it is just as well that they were suffering from impotence and could not have been cured by the treatment. Nevertheless, their case histories would now make fascinating reading with the wisdom of hindsight.

R. V. Short

Department of Perinatal Medicine,
Royal Women's Hospital, University of Melbourne,
Melbourne 3053, Australia

1. Gao, F. *et al.* *Nature* **397**, 436–441 (1999).

2. Short, R. V. *Acta Paediatr. Suppl.* **422**, 3–7 (1997).

3. Gosden, R. *Cheating Time: Science, Sex and Ageing* (Macmillan, London, 1996).

4. Voronoff, S. *Scientific American*, October, 226–227 (1925).

Three planets for Upsilon Andromedae

Jack J. Lissauer

In 1997, a planet of roughly Jupiter's mass was discovered in orbit around the star υ Andromedae. Two more have now been identified.

In the 1990s, the quest to find planets outside our own Solar System finally succeeded. The first extrasolar planets to be identified are in orbit around a pulsar¹, a rapidly rotating neutron star which is quite different from our own familiar Sun. The first planetary companion to a main-sequence star — a hydrogen-burning star like the Sun — was discovered in 1995, using Doppler measurements of the reflex motion (wobble) that the planet's orbit induces in the star 51 Pegasi². Nearly 20 more main-sequence stars have since been found to possess a single companion of roughly Jupiter mass³. But now we have news of something else again — the discovery, announced last week by Paul Butler and his colleagues⁴, of a system of three massive planets orbiting the star υ Andromedae.

All of the planets known to orbit main-sequence stars other than the Sun have been identified using the Doppler technique, which measures changes in the star's velocity component along the line of sight. These observations can be inferred as perturbation of the star by a companion planet. They yield the planet's orbital period and orbital eccentricity, and the product of its mass and the sine of the inclination of its orbital plane to the plane of the sky, $M\sin(i)$. Because $\sin(i)$ cannot exceed unity, the measured radial-velocity variations of the star provide a lower limit to the planet's mass.

Determining the masses and orbits of three planets travelling about a star using Doppler data is a highly complicated business, because the data have significant intrinsic scatter and the signatures of the planets interfere with one another. Butler *et al.*⁴ looked at two sets of data, obtained respectively at the Lick Observatory in California and by the Advanced Fibre-Optic Echelle (AFOE) spectrometer at the Whipple Telescope in Arizona. They analysed these data independently, and when combined, to derive three sets of planetary parameters. The similarity of the values of these parameters (see Fig. 1) lends confidence to the three-planet solution.

The star υ Andromedae has a mass about 30% greater than that of our Sun, is about halfway through its six-billion-year main-sequence lifetime and is about 44 light years from our Solar System. The innermost planet was first discovered in 1997 (ref. 5). It has a nearly circular orbit with a period of 4.6 days,

and a minimum mass about 0.7 times that of Jupiter; these properties are similar to those of the only known planet of 51 Pegasi and to other star-grazing planets.

What about the two newcomers identified by Butler *et al.*? The middle planet has a period of about 242 days, a mass about twice that of Jupiter and an orbital eccentricity, e , of approximately 0.22 (comparable to the eccentricities of Mercury and Pluto, and larger than the orbital eccentricities of all other planets known in our Solar System). The orbital period of the outer planet is $\sim 1,270$ days; $e \sim 0.36$; and mass ~ 4.1 times that of Jupiter. These latter values are not well constrained, because most of the Doppler data were taken within a 6.5-year interval, less than two orbital periods of the planet. But the masses and eccentricities of the outer

two bodies are typical of half-a-dozen known extrasolar giant planets with orbital periods of around a year.

Additional constraints on the υ Andromedae planets come from astrometric and photometric data, as well as from numerical orbit integrations. The Hipparcos astrometric satellite observed the position of υ Andromedae on the plane of the sky several times during the early 1990s. No wobble is detectable in these data, which implies that the outermost planet is less than ten and probably no more than five Jupiter masses. Photometric measurements of the brightness of υ Andromedae around the times that the inner planet would pass in front of (transit) the star if its orbital plane were nearly along the line of sight ($\sin(i)$ close to 1) rule out transits deeper than about 0.0002 to 0.0003 magnitude (G. W. Henry, personal communication). This means that the inclination of the planetary orbit must be less than 83° or that we are dealing with an extremely dense, exotic object.

As far as numerical integrations are concerned, the preliminary results of calculations by myself and a colleague, E. J. Rivera, imply that a system having the parameters derived from the AFOE data only, with $\sin(i)$ of approximately 1, would tear itself apart within a million years or so (Fig. 1). (The chaotic nature of such a trajectory precludes

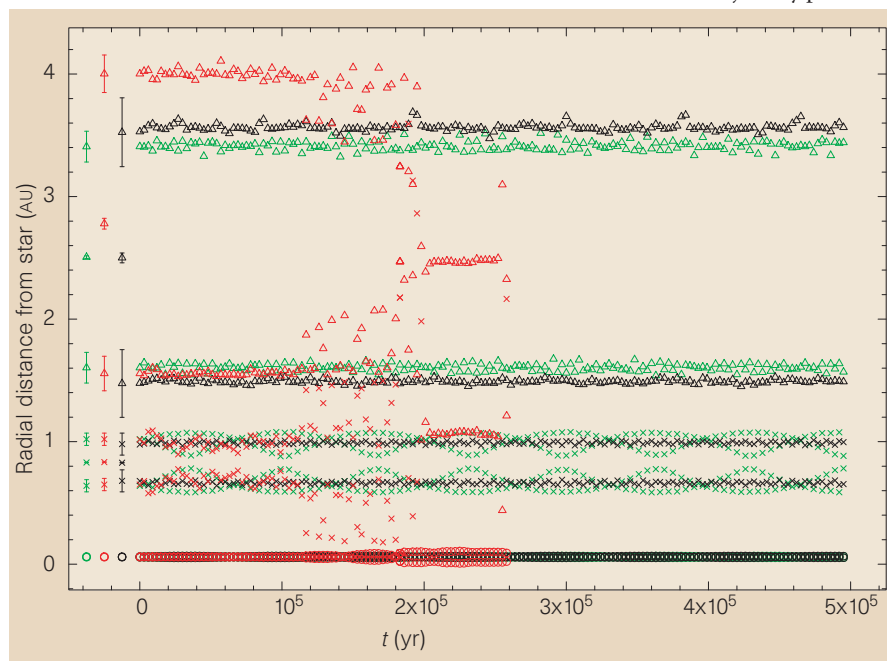


Figure 1 The reported orbital parameters for the planets around υ Andromedae⁴ are given on the left side of the plot. The orbit of each planet is represented by three points — the periastron and apoastron (closest and farthest distances from the star, bottom and top points) and the semimajor axis (roughly, the average orbital distance). Triangles represent the outer planet, crosses the middle planet and circles the inner planet. Error bars are shown for the outer two planets; observational uncertainty is far less for the innermost planet. Data from the Lick Observatory are given in green and from the AFOE (Whipple Telescope) in red; combined data are in black. The rest of the plot shows the temporal evolution of the periastra and apostra of the three planets using the nominal parameters on the left as initial conditions. The systems using the combined and Lick data sets continue to be stable for the entire ten-million-year integration interval. (Calculations and figure by E. J. Rivera of SUNY, Stony Brook, and J.J.L. using the SwIFT symplectic MVS package⁸.)

a precise calculation.) The star, and thus presumably its planets, are about three billion years old, and it would be extremely unlikely that the system just happens to be so near to the end of its expected lifetime. What the calculation implies, then, is that the nominal AFOE parameters are unlikely to be correct. Nonetheless, the characteristics of the system could well be within the error bars quoted for the AFOE data.

By contrast, systems using the Lick or combined Lick/AFOE data sets appear to be stable for at least ten million years and possibly much longer. Additional integrations, not shown in Fig. 1, imply that systems with more massive planets (smaller $\sin(i)$) are generally less stable, especially if the orbits are substantially inclined to one another, and that changing the orbital period of the outer planet within the range allowed by the data can have a big effect on the stability of the system.

Neither the star ν Andromedae nor any of its three planets appear in themselves to be particularly unusual, but the system as a whole stands out in the currently known menagerie of extrasolar planets. Its existence shows that several massive planets can orbit far closer to a star than do the gas giants (Jupiter and Saturn) and smaller ice giants (Uranus and Neptune) of our Solar System. So if giant planets all form far enough out in a protoplanetary disk for ice to condense (the conventional, but not exclusive, view of theorists⁶), then the migration or mutual scattering (or both) that brings them in towards their star is not always violent enough to

destroy themselves or their brethren.

The ν Andromedae system is just one example of a wide variety of planetary configurations that can be expected to exist in our Galaxy. All of the extrasolar planets thus far discovered orbiting main-sequence stars are more massive than Saturn, and most either orbit very close to their stars or travel on much more eccentric paths than do any of the major planets in our Solar System. The Sun's planets are all either low in mass or travel on distant orbits from their star, and they are therefore more difficult to discover using the Doppler technique. So it could be that most planetary systems will turn out to be like ours.

There is much, much more to come. With the projected advances in detection technology, including schemes to find Earth-like planets⁷, we are at the beginning of a Golden Age of extrasolar planetary studies. $\square$

Jack J. Lissauer is in the Space Science Division, NASA Ames Research Center, Moffett Field, California 94035, USA.

e-mail: JLISSAUER@ringside.arc.nasa.gov

1. Wolszczan, A. & Frail, D. A. *Nature* **355**, 145–147 (1992).
2. Mayor, M. & Queloz, D. *Nature* **378**, 355–359 (1995).
3. Marcy, G. W. & Butler, R. P. *Annu. Rev. Astron. Astrophys.* **36**, 57–97 (1998); <http://www.physics.sfsu.edu/~gmarcy/planetsearch/planetsearch.html>
4. Butler, R. P. *et al. Astrophys. J.* (submitted); <http://www.physics.sfsu.edu/~gmarcy/planetsearch/upsand/upsand.html>; <http://cfa-www.harvard.edu/afoe/>
5. Butler, R. P., Marcy, G. W., Williams, E., Hauser, H. & Shirts, P. *Astrophys. J.* **474**, L115–L118 (1997).
6. Wuchterl, G., Guillot, T. & Lissauer, J. J. *Protostars and Planets IV* (eds Mannings, V., Russel, S. & Boss, A. P.) (Univ. Arizona Press, Tucson, in the press).
7. <http://www.kepler.arc.nasa.gov/>; <http://ast.star.rl.ac.uk:80/darwin/>
8. Levison, H. F. & Duncan, M. J. *Icarus* **108**, 18–29 (1994).

ions of the salt and the molecular pore in the α -haemolysin channel. When a bias is applied across the membrane, the current flow is tiny (at the picoampere level), but it can nonetheless be measured with modern room-temperature semiconductor electronics.

The innovative step here¹ is to sensitize the nanopore to specific organic chemical species by using β -cyclodextrin as an adapter. β -cyclodextrin, a doughnut-like molecule made of seven sugar units, diffuses into the α -haemolysin channel, partially obstructing its water-filled pore. The authors show that the cyclodextrin molecule is exquisitely sensitive to different guest molecules. Thus, when bound to cyclodextrin, members of the adamantane family of petroleum derivatives can be distinguished, as can members of the group of tricyclic pharmaceuticals that include imipramine and promethazine. These guest molecules each bind to β -cyclodextrin for milliseconds and make their presence known by altering the electrical conductivity of the α -haemolysin pore in which the β -cyclodextrin resides.

In neurobiology and biophysics, it is routine to measure picoampere ionic currents passing through single protein channels in biological membranes or planar lipid bilayers³. Such measurements are usually made to determine the properties and behaviour of the channel, but they can also reveal details about the concentration and other properties of molecules that are able to pass through or into the channel. Very small alterations in the distribution of charges lining the pore can give rise to relatively large changes in the ionic flux, so a membrane channel can serve as an amplifying transducer. Such a transducer allows one to detect extremely small modulations ($<0.02 kT$) in the energy barrier to ionic transport⁴, and does so at bandwidths that permit high-speed ($<50 \mu\text{sec}$) measurements. Thus, membrane channels have been used to effect high-speed molecular counting and sizing^{5,6}, and show promise for high-speed sequencing of polynucleotides⁷. Channels have also been modified to incorporate design elements that convert them into sensitive, analyte-driven switches⁸, or genetically engineered to create selective, divalent metal sensors⁹.

Bayley and his colleagues have long puzzled over how to engineer or redesign the α -haemolysin channel as a biosensor that can discriminate between different organic molecules⁹. They reasoned that the sensitivity and extraordinarily rich informational content of single-channel recordings reduce the requirement for the channel to contain a highly selective binding site. By fashioning a single-channel detector, rather than one that integrates signals from numerous sensor molecules, several independent

Biochemical sensors

Adapting to nanoscale events

Daniel Branton and Jene Golovchenko

A hallmark of twentieth-century science has been the continual development of experimental strategies to observe individual atomic-scale 'events'. These strategies ultimately rely on significantly amplifying the consequences of a selective microscopic interaction, for example the chemical development of a silver halide grain in a photographic emulsion, the condensation of a droplet around a single ion in a cloud chamber, or the charge amplification in electron multiplier devices.

New strategies for detecting and characterizing single-molecule events are now emerging in the biochemical sciences. The latest example comes from Gu *et al.* on page 686 of this issue¹. This group, working in Hagan Bayley's laboratory, show how measurement of ionic transport through a single, atomic-scale pore in an insulating membrane can be used to detect organic molecules of relative molecular mass as low as 100. Coupled with highly sensitive semi-

conductor electronics, the membrane-pore system can amplify these currents on timescales commensurate with the interaction times between the molecule and the pore. The system thus reveals the presence, nature and interaction of single molecules with the pore. Remarkably, a single protein channel can be adapted for simultaneous analysis of a mixture of organic molecules.

Gu *et al.*¹ assemble the nanoscale chemical and mechanical building blocks of their detector from the toolbox of biochemistry; the key components are shown in Figs 1 and 2 of the paper on page 687. Seven α -haemolysin molecules (a bacterial toxin) self-assemble to form a channel with a 15-Å-diameter aqueous pore through a 50-Å-thick lipid bilayer membrane². The membrane separates two chambers filled with a conducting salt solution. Because the lipid bilayer is a nearly perfect insulator, the d.c. electrical conductance across the membrane is determined by the interaction between

features can be simultaneously recorded for each analyte-binding event. For example, the current amplitude, voltage-dependence and typical kinetics can, together, produce a distinctive signature for each of several analytes, even though the sensor binding selectivity is minimal. Furthermore, use of an adapter together with an information-rich single-channel recording method may avoid the contradictory requirements of high selectivity (to achieve specificity) but rapid reversibility (to achieve speed in a changing environment). Because fresh adapters can be continuously flushed into the device, even an adapter (or carrier) that binds tightly to the analyte can maintain the system's speed if the adapter association with the channel exhibits rapid reversibility.

Although it will obviously be of great interest to identify the atomic-level changes that account for changes in ionic current flow upon adapter and analyte binding, there remain many questions about the ultimate capabilities of single-channel pores in atomic-scale investigations of individual molecules. What are the fundamental physical processes that limit the sensitivity and speed with which single-molecule measurements can be made? Can electronic currents be used to replace ionic currents? Can robust

membrane-pore systems be engineered using advanced materials science and nanolithography techniques, perhaps in conjunction with the methods of biological or chemical engineering? Can sensitive electronic circuitry and photonic sensing capabilities ultimately be integrated directly into a membrane-pore system, as they have for other chemical sensors^{8,10}? These questions may not be answerable today. But, in the twenty-first century, exciting new nanoscale tools and molecular machines will surely emerge at this interface of biology, chemistry and physics. □

Daniel Branton and Jene Golovchenko are in the Department of Molecular and Cellular Biology and the Department of Physics, Harvard University, Cambridge, Massachusetts 02138, USA.
e-mail: dbranton@harvard.edu

1. Gu, L.-Q., Braha, O., Conlan, S., Cheley, S. & Bayley, H. *Nature* **398**, 686–690 (1999).
2. Gouaux, E. *J. Struct. Biol.* **121**, 110–122 (1998).
3. Hammill, O. P. et al. *Pflügers Arch.* **391**, 85–100 (1981).
4. Sigworth, F. J. *Biophys. J.* **47**, 709–720 (1985).
5. Bezrukov, S. M., Vodyanoy, I. & Parsegian, V. A. *Nature* **370**, 279–281 (1994).
6. Kasianowicz, J. J., Brandin, E., Branton, D. & Deamer, D. W. *Proc. Natl Acad. Sci. USA* **93**, 13770–13773 (1996).
7. Akeson, M. et al. *Biophys. J.* **76**, A152 (1999).
8. Cornell, B. A. et al. *Nature* **387**, 580–583 (1997).
9. Braha, O. et al. *Chem. Biol.* **4**, 497–505 (1997).
10. Ricco, A. J. & Crooks, R. M. *Acc. Chem. Res.* **31**, 219–227 (1998).

Neurobiology

Attraction is relative not absolute

Masataka Watanabe

Suppose you like wine very much, and are faced with the choice of ordinary table wine or Beaujolais Nouveau — the latter is probably very attractive. But if presented with Beaujolais Nouveau and a superior wine, such as Romanée-Conti, the attraction of the Beaujolais Nouveau may subside. A similar phenomenon called 'reinforcement contrast effect' seems to occur in animal-learning studies, and Tremblay and Schultz¹ now show (page 704 of this issue) that a region of the brain called the orbitofrontal cortex may be actively involved in this process.

Fifty years ago, in an experiment by Zeaman², rats were trained to run to a goal, for which they received a large reward. They were then shifted suddenly to a small reward. A second group of rats was trained the other way round; that is, they were first given a small reward and then shifted to a larger one. The animals that were shifted from a small to a larger reward immediately ran faster than would have been expected if the large reward were used alone. Conversely, those rats shifted from a large to a small reward ran more slowly than would have been expected with the small reward alone. The results indicated that the relative — but not absolute — magnitude of reward deter-

mines the attraction, as assessed by the eagerness of the rats to run to the goal.

Where in the brain is the attraction of a reward evaluated? To find out, Tremblay and Schultz¹ trained monkeys in a modified delayed-response task (Fig. 1). When the animal pressed a lever at the bottom of an experimental panel, a picture 'cue' was presented briefly on either the right or left side of the panel. Then there was a delay, after which the animal had to press the lever (right or left) above which the picture had been presented. This meant that, to obtain a reward, the monkey had to retain a memory of where the picture (although not what

picture) was presented. Within a block of several tens of trials, the same two pictures were used continuously, and each picture always predicted one kind of reward.

Tremblay and Schultz found that whereas neurons in the lateral part of the prefrontal cortex showed changes in activity during the delay period, according to the position or identity of the picture, neurons in the orbitofrontal cortex were not sensitive to these factors. The orbitofrontal cortex — so called because it is located above the orbit, or eye socket — occupies the ventral surface of the prefrontal cortex (Fig. 2, overleaf). Humans and monkeys with damage to the orbitofrontal cortex often show impaired motivational and emotional behaviour³, such as altered reward preferences⁴. Consistent with this, the authors found that many orbitofrontal neurons were sensitive to the different rewards. They observed such responses when the picture cue was presented, during the expectation of reward, or after the monkey had received the reward.

Figure 1 shows the activity of such an orbitofrontal neuron. The neuron is more active when the monkey sees a square, which predicts a morsel of apple, than when it is presented with a triangle, which predicts a piece of lettuce (Fig. 1a). But in the next trial block, when a square predicts apple and a plus sign a piece of banana, the same neuron shows higher activity to the cue for banana, because monkeys prefer bananas to apples, and apples to lettuce (Fig. 1b). The attraction of the apple now seems to subside, and the neural activity to the square cue is reduced. In fact, it is similar to the activity for the triangle cue, which predicted the lettuce reward in the previous trial block. So, the activity of orbitofrontal neurons reflects the relative, but not absolute, attraction of the reward.

Once a monkey has had enough of the reward, the orbitofrontal neurons stop responding to the reward and to the cues that predict it³. This process is specific to the nature of the reward, indicating that activity of the orbitofrontal neurons is determined by how much the monkey wants the reward³. The work of Tremblay

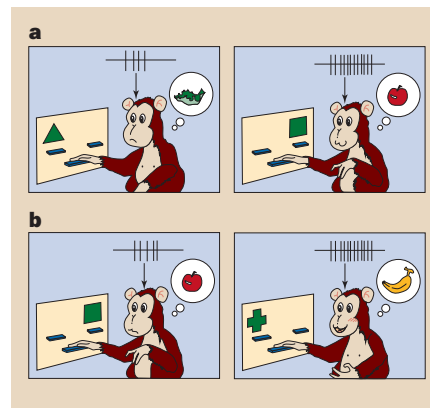


Figure 1 Activity of an imaginary orbitofrontal neuron reflecting relative reward preference. To gain a reward, the monkey must remember where (right or left side) the cue is presented. a, A triangle predicts a lettuce reward, and a square predicts an apple. The monkey prefers apple to lettuce, so activity of the neuron is greater when the animal sees the square. b, This time, a square predicts an apple and a plus sign predicts a banana. Because the monkey prefers banana to apple, activity of the neuron is greatest with the plus sign, and much lower than in the previous test with the square. (Figure illustrated by Chiharu Tomita.)

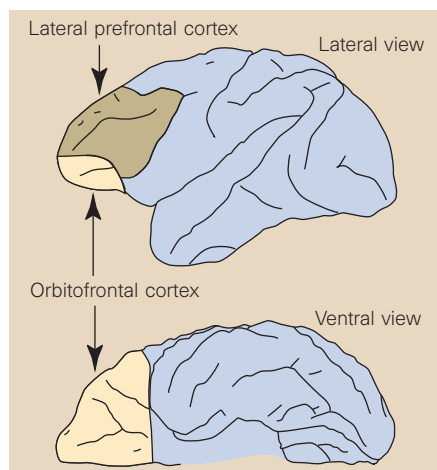


Figure 2 Lateral and ventral views of the orbitofrontal cortex of the monkey. Tremblay and Schultz¹ show that this region of the brain is involved in assessing the relative — but not the absolute — attraction of a reward.

and Schultz¹ adds a new dimension to characterization of the orbitofrontal neuronal activity — the relative attraction of the incentive.

It is important to place these results into the broader context of the part played by the prefrontal cortex in cognition. Studies of the prefrontal cortex have focused on the lateral prefrontal cortex in relation to working memory^{5,6}. Current debates centre on whether dorsal and ventral parts of the lateral prefrontal cortex are functionally segregated according to what information

is retained in working memory, or how that information is kept there^{5,6}. But the orbitofrontal cortex — which occupies quite a large proportion of the prefrontal cortex — is not essential for working memory. If this area of the brain is damaged, working-memory tasks are not impaired. Moreover, clinical and noninvasive studies indicate that the orbitofrontal cortex is important for estimating the pleasantness or value of objects, emotional reactions^{3,7} and motivational operations such as expectation of reward^{8,9}. The lateral prefrontal cortex, which is considered to be more involved in cognitive operations, is also implicated in reward expectancy^{8,9}. It is clear, then, that theories to describe the functions of the primate prefrontal cortex will have to include the motivational parameters that underlie much of our behaviour¹⁰. □

Masataka Watanabe is in the Department of Psychology, Tokyo Metropolitan Institute for Neuroscience, Musashidai 2-6, Fuchu, Tokyo, 183-0042, Japan.
e-mail: masataka@tmin.ac.jp

1. Tremblay, L. & Schultz, W. *Nature* **398**, 704–708 (1999).
2. Zeaman, D. J. *Exp. Psychol.* **39**, 466–483 (1949).
3. Rolls, E. T. *The Brain and Emotion* (Oxford Univ. Press, 1999).
4. Baylis, L. L. & Gaffan, D. *Exp. Brain Res.* **86**, 617–622 (1991).
5. Goldman-Rakic, P. S. *Phil. Trans. R. Soc. Lond. B* **351**, 1445–1453 (1996).
6. Petrides, M. *Phil. Trans. R. Soc. Lond. B* **351**, 1455–1462 (1996).
7. Damasio, A. *Descartes' Error* (Putnam, New York, 1994).
8. Thut, G. *et al. NeuroReport* **8**, 1225–1228 (1996).
9. Watanabe, M. *Nature* **382**, 629–632 (1996).
10. Watanabe, M. *Rev. Neurosci.* **9**, 225–241 (1998).

Atmospheric chemistry

Uncertain road to ozone recovery

Paul J. Fraser and Michael J. Prather

The manufacture and emission of chlorofluorocarbons (CFCs), our inadvertent global experiment in modifying the Earth's stratosphere, has damaged the ozone layer for decades to come. The Montreal Protocol, which was agreed in 1987 and revised several times in the 1990s, and has the aim of reducing global emissions of ozone-depleting chemicals, has gone a significant way to solving the 'ozone problem'¹. However, as described by Montzka and colleagues on page 690 of this issue², the road to ozone recovery remains uncertain.

The authors report the good news that, by 1997, the lower-atmospheric burden of ozone-depleting halogen compounds (chloro- and bromocarbons) had declined by 3% from its 1993–94 peak; and they provide a reminder that this decline is almost entirely due to the rapid atmospheric removal of the solvent methyl chloroform (CH_3CCl_3), global emissions of which are now virtually zero. But concentrations of other halocarbons — carbon tetrachloride

(CCl_4) and CFC-11 (CCl_3F), for instance — are falling more slowly, and those of others such as CFC-12 (CCl_2F_2) are still rising. If the decline in emissions of these other halocarbon species does not gather pace, the overall fall in halogen concentrations in the atmosphere will stall sometime during the next decade.

What are the barriers to ozone recovery? Montzka *et al.* have identified several, namely, emissions of refrigerant CFCs, of HCFCs (hydrochlorofluorocarbons) as interim CFC replacements, and of halons (bromocarbons) as fire-fighting chemicals. In the developed world, a large 'bank' of CFC-12 in old refrigeration and vehicle air-conditioning systems continues to leak slowly into the atmosphere. Although new CFC production declined by a factor of 4–5 from 1986 to 1995 in accord with the Montreal Protocol, the developing world (China, India, Mexico) has increased production by 2–3 fold (largely of CFC-12), so that in 1995 it accounted for 45% of global CFC production³. This figure

is presumably even larger now. The combined effect of these emissions is continuing growth of CFC-12 in the atmosphere (Fig. 1a, overleaf), and this is the biggest single long-term threat to ozone recovery.

Fortunately, HCFC emissions to date have been only 50% of those permitted under the Protocol; in large part this is because HCFCs are due to be phased out by 2030, and some parts of the refrigeration and air-conditioning industry have hesitated to invest in short-term HCFC technologies. Likewise, HCFCs have not replaced CFCs in aerosols or foam plastics, where hydrocarbons or 'not-in-kind' substitutes have filled these niches. Yet HCFC use — and therefore emissions and atmospheric concentrations — are predicted to increase substantially, and legally, under the Protocol over the next decade⁴ as HCFCs continue to be used as CFC substitutes in refrigeration (Fig. 1b).

Another emerging player is halon-1211 (CBrClF_2), a fire-fighting chemical used increasingly in developing countries such as China and Korea, which now account for over 95% of global production of halons³. Under the Protocol, emissions of halon-1211 were thought to have peaked in the late 1980s (ref. 4), but atmospheric observations indicate near constant^{2,5} or possibly growing⁶ emissions (Fig. 1c and d). Using inventories based on atmospheric observations, emissions in the late 1990s are 50–60% higher than those predicted from global production data⁶. Presumably global production has been underestimated, or emissions from the bank are larger than assumed (or both). Studies by Montzka *et al.*² and a group involving one of us⁶ conclude that, in the near term (as opposed to CFC-12 in the long term), halon-1211 emissions pose the largest threat to ozone recovery.

Prompt and continuing progress towards ozone recovery requires that emissions of halocarbons must be reduced faster than is apparent from current observations of the atmosphere. The United Nations Environmental Programme is facilitating such an accelerated phasing out of halon use in China, with production of halon-1211 scheduled to end by 2006, four years ahead of the Protocol⁷. In addition, the technology exists to recycle or destroy the bank of CFC refrigerants that would otherwise vent to the atmosphere when refrigeration systems are scrapped. The uses of HCFCs could be restricted to refrigeration, where their identification and recycling is again possible, and emphasis could be placed on use of very short-lived gases such as HCFC-123 (CHCl_2CF_3) that pose a minor risk to the ozone layer⁸. All in all, it is possible that ozone depletion can be halted in the next decade. But it will require a level of global stewardship that still poses a substantial challenge to all parties to the Montreal Protocol.

NATURE

100 YEARS AGO

The Groundwork of Science; a Study of Epistemology. By St. George Mivart. The chief definite conclusions which are drawn are that it (the universe) cannot consist of one kind of energy only, that it is impossible that intellect can have been evolved from mere physical force, and that animals show no signs of latent intellectuality. It is further insisted "that the portion of truth which we are able to attain to in our investigations of the cosmos, is but an unimaginably small portion of the whole"; a statement which will, we imagine, not be seriously challenged by workers in science. To the latter, viz. the science workers, Dr. Mivart devotes some attention in the concluding pages of his book. The narrowing effect of extreme specialism upon the mind is an undoubted evil ... But there is the opposite evil of becoming diffuse to the extent of a practically useless attenuation of the mental faculties.

From *Nature* 20 April 1899.

50 YEARS AGO

Mr. H. E. Hadley, well known as the author of many elementary text-books of physics, died on March 6, at the age of eighty-two. Mr. Hadley had lived rather a retired life. He was appointed headmaster of a small science school in Kidderminster, where he combined with physics a lectureship in chemistry. In those days, physics was considered of less importance than chemistry; but Mr. Hadley was always at heart a physicist. ... Mr. Hadley was a contemporary of Sir Richard Gregory and H. G. Wells at [The Royal College of Science, London], and from his training there and association with C. V. Boys he acquired a special genius for making his physical apparatus. His teaching equipment at Kidderminster was largely of his own making, and it would seem that this gave physics an added attraction to his students.

We regret to announce the following death: Mr. Will Hay, well known as an actor and also a distinguished amateur astronomer, on April 18, aged sixty.

From *Nature* 23 April 1949.

Many more extracts like these can be found in *A Bedside Nature: Genius and Eccentricity in Science, 1869–1953*, a 266-page book edited by Walter Gratzer. Contact Lisa O'Rourke. e-mail: l.ourourke@nature.com

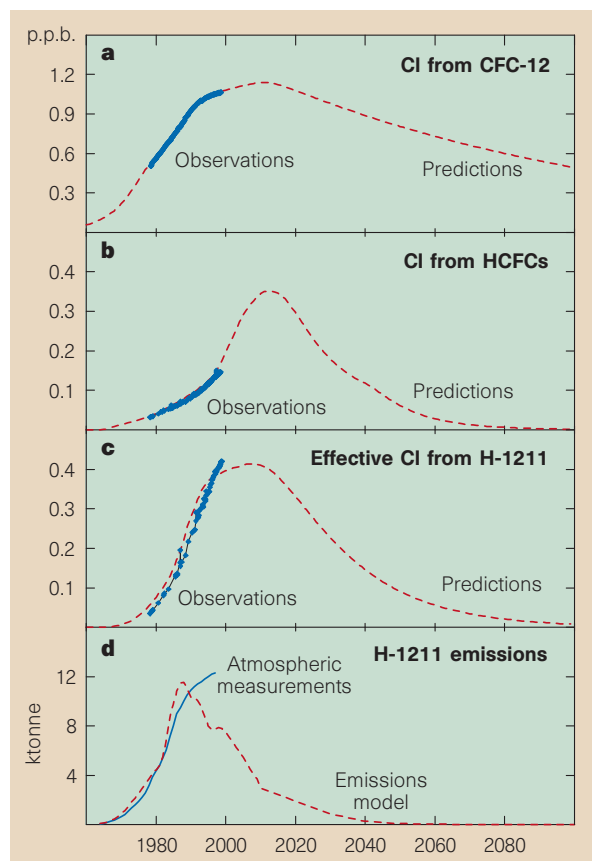


Figure 1 Ozone-depleting trace gases now and in the future — observed^{6,15,16} and predicted⁴ atmospheric chlorine loading, in parts per billion of Cl, from various halocarbons. a, CFC-12; b, HCFCs; c, halon-1211 (effective Cl). d, Halon-1211 emissions calculated from atmospheric measurements and from a production-based emissions model⁶.

Ozone depletion is the dark side of the CFC experiment. But, for science, there is a bright side — the great progress in atmospheric chemistry, made over the past 25 years, which stems in part from the scientific and political drive to understand CFCs and ozone depletion, and in part from the study of the rapid yet transient rise of the synthetic halocarbons. For example, the earliest measurements of CFC-11 envisaged its use as a tracer of air motions⁹, and it has become a standard test of global models of atmospheric chemistry¹⁰ (and also a diagnostic indicator of ocean circulation); the first meaningful measure of the mean tropospheric concentration of hydroxyl radicals (OH), which destroy CH_3CCl_3 and HCFCs, is derived from observations of CH_3CCl_3 (refs 11, 12); CFC pollution events in Ireland are used to calibrate European emissions of other greenhouse gases¹³; and study of CH_3Br (a crop and soil fumigant) has shown us the importance of the ocean in determining the atmospheric residence time of a soluble trace gas, and has resulted in revision of the concept of ozone-depletion potential for short-lived gases¹⁴.

What new opportunities will the second phase of the CFC experiment provide? Through its latitudinal gradient, the decay of CH_3CCl_3 will give us a measure of the north–south hemispheric difference in OH concentrations, plus a long-term record of possible OH trends. And, after sources of CFC-11 have diminished sufficiently, its atmospheric decay will provide us with the

first accurate lifetime for a CFC, and the north–south differences at the surface will measure hemispheric asymmetry in stratosphere–troposphere exchange. It is serendipitous indeed that CFCs and related halocarbons have provided, and will continue to provide, some of the benchmarks of progress in atmospheric chemistry. □

Paul J. Fraser is at CSIRO Atmospheric Research/CRC Southern Hemisphere Meteorology, PMB 1, Aspendale, Victoria 3195, Australia.

e-mail: paul.fraser@dar.csiro.au

Michael J. Prather is in the Earth System Science Department, University of California, Irvine, California 92697-3100, USA.

e-mail: mprather@uci.edu

1. Prather, M. *et al.* *Nature* **381**, 551–554 (1996).
2. Montzka, S. A. *et al.* *Nature* **398**, 690–694 (1999).
3. Oberthur, S. *Production and Consumption of Ozone-Depleting Substances 1986–1995* (Deutsche Gesellschaft für Technische Zusammenarbeit, Berlin, 1997).
4. Velders, G. J. M. & Madronich, S. in *Scientific Assessment of Ozone Depletion: 1998* Ch. 11 (UNEP/WMO, in the press).
5. Butler, J. H. *et al.* *J. Geophys. Res.* **103**, 1503–1511 (1998).
6. Fraser, P. J. *et al.* *J. Geophys. Res.* (in the press).
7. Executive Committee of the Multilateral Fund for the Implementation of the Montreal Protocol *China's Halon Sector Strategy* (UNEP/OzL.Pro/ExCom/23/11, 21 October, 1997).
8. Wuebbles, D. & Calm, J. *Science* **278**, 1090–1091 (1997).
9. Lovelock, J. E. *Nature* **230**, 379–381 (1971).
10. Prather, M. *et al.* *J. Geophys. Res.* **92**, 6579–6613 (1987).
11. Lovelock, J. E. *Nature* **267**, 32 (1977).
12. Prinn, R. *et al.* *Science* **269**, 187–192 (1995).
13. Simmonds, P. G. *et al.* *Atmos. Environ.* **30**, 4041–4063 (1996).
14. Butler, J. H. *Geophys. Res. Lett.* **21**, 185–189 (1994).
15. Prinn, R. *et al.* in *Scientific Assessment of Ozone Depletion: 1998* Ch. 1 (UNEP/WMO, in the press).
16. Kurylo, M. *et al.* in *Scientific Assessment of Ozone Depletion: 1998* Ch. 2 (UNEP/WMO, in the press).

DNA repair

Gatekeepers of recombination

James E. Haber

Double-strand breaks on chromosomes may arise during DNA replication, by enzymatic cleavage, or from ionizing radiation. This damage can be repaired in several ways. The two broken ends may simply be stuck back together by a process called nonhomologous end-joining, or the break may be repaired through genetic exchange with a homologous chromosome (homologous recombination). How do cells choose which pathway to use? On page 728 of this issue, Van Dyck *et al.*¹ suggest that the Rad52 protein may play the role of 'gatekeeper' in homologous recombination, just as the Ku proteins seem to do for nonhomologous end-joining.

Homologous recombinational repair usually proceeds by gene conversion, in which new DNA is copied from a donor template (Fig. 1a). This process begins by strand invasion of the donor sequence by the 3' ends of the double-strand break. These ends are exposed when an exonuclease enzyme nibbles back (resects) the 5' strands. The 3' end makes extensive base pairs with the complementary donor strand and is then used as the primer by a DNA polymerase to initiate new DNA synthesis to repair the break. In the yeast *Saccharomyces cerevisiae*, homologous recombination involves replication protein A (RPA)², which binds the single-stranded DNA, the recombination proteins encoded by the *RAD51*, *RAD52*, *RAD54*, *RAD55* and

RAD57 genes³, and, apparently, all three of the replicative DNA polymerases with their associated factors⁴.

So far as we can tell, the basic features of recombination are conserved from bacteria, through yeast, to humans. The key step — that of strand exchange, when the 3' end invades the donor duplex — is carried out within a filament of Rad51, whose bacterial homologue is RecA⁵. Rad52 seems to occur only in eukaryotes. But surprisingly, in *Saccharomyces* at least, it is the key protein for recombination. Almost all homologous recombination, whether spontaneous or induced by double-strand breaks, is eliminated in yeast *rad52* mutants. Curiously, Rad51-deficient cells can still carry out several types of homologous recombination and repair, including those initiated by double-strand breaks^{6,7}. The situation is not yet clear for higher eukaryotes. In vertebrates, a lack of Rad52 is not nearly as serious as the absence of Rad51, which is lethal⁸. Possibly a Rad52 homologue will be found. On the other hand, the apparent absence of a Rad52 homologue in the completely sequenced nematode genome should make us pause in our quest for some unity of mechanism (although I will take modest wagers that such a gene may yet turn up).

By itself, Rad52 stimulates annealing between complementary regions of single-stranded DNA^{9,10}. This activity could explain

why Rad52 is, so far, the only recombination protein required for the alternative homologous repair pathway, single-strand annealing (Fig. 1b). Genetic and biochemical experiments have also shown that Rad52 interacts with both Rad51 and RPA, and that it stimulates Rad51-mediated strand-exchange reactions *in vitro*^{11,12}. The strand-exchange reaction mediated by Rad51 is much more dependent on a precise stoichiometry of interacting molecules and on the involvement of auxiliary proteins than is the equivalent RecA assay. Rad52 itself forms ring structures in the absence of DNA¹³.

Now, Van Dyck *et al.*¹ provide electron microscopic evidence that human Rad52 — apparently as multimeric rings — selectively binds to DNA ends. They also show biochemically that Rad52 protects DNA ends from exonuclease attack. The authors propose that Rad52 could help to load Rad51 onto a single-stranded DNA end. This suggestion is consistent with the elegant cytological observations of Gasior *et al.*¹⁴, who showed that the formation of Rad51 filaments during recombination in meiotic yeast cells is blocked in the absence of Rad52.

Van Dyck *et al.* also provide compelling evidence that purified Rad52 strongly promotes end-to-end joining of DNA — both between and within molecules. The end-binding and end-joining properties of Rad52 are similar to what has been shown for Ku proteins *in vitro*. The end-binding property could explain why, *in vivo*, the Ku proteins partially protect DNA ends from digestion by 5' to 3' exonucleases⁴. Ku proteins are essential for nonhomologous end-joining (Fig. 1c). In end-joining, junctions are formed between DNA ends, sometimes precisely and sometimes with the loss of hundreds of nucleotides. Most junctions show that at least one base pair has been formed between overhanging single strands of resected DNA ends¹⁵. Seven proteins are known to be required for this process — Ku70 and Ku80, DNA ligase IV and the associated XRCC4 protein, and Mre11, Rad50 and Xrs2/p95 — and these same proteins are also implicated in the end-joining processes of V(D)J recombination in the immune system. The idea that, *in vitro*, Rad52 also promotes end-joining in the presence of DNA ligase is fascinating, but it is unclear what its relevance is *in vivo*, given that it does not seem to be involved in nonhomologous end-joining.

Van Dyck *et al.* propose that Rad52 and Ku are competing agents, channelling the ends of a double-strand break into alternative repair pathways. Which pathway a cell selects has great significance. End-joining events are generally mutagenic at the site of the double-strand break. Repair by gene conversion, on the other hand, should accurately repair the break, although it can lead to chromosome rearrangements. The contest

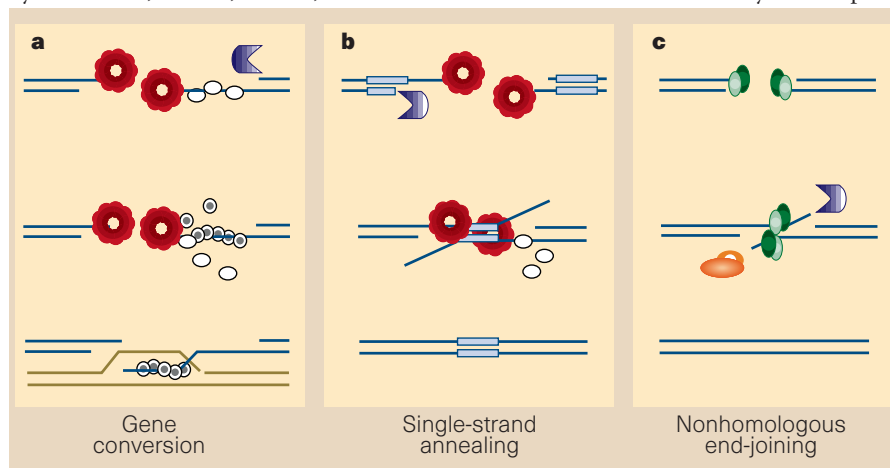


Figure 1 Alternative pathways of double-strand break repair. **a**, Gene conversion. This process depends on Rad52 (red), which interacts with the single-stranded DNA binding complex RPA (white), and with the Rad51 strand-exchange protein (grey). Rad52 may facilitate the displacement of RPA by Rad51. **b**, Single-strand annealing. DNA ends are extensively resected, in part by the nuclease complex Mre11–Rad50–Xrs2 (purple), until substantial regions of homology flanking the break are exposed. This process depends on Rad52, consistent with its role as a strand-annealing protein. The association of Rad52 rings with each other may help bring ends together to begin the search for homology. **c**, Nonhomologous end-joining. The Ku70–Ku80 proteins (green) bind to, and facilitate joining of, DNA ends. Mre11–Rad50–p95, and DNA ligase IV with its associated XRCC4 (orange) are required, but Rad52 is not. (Single-strand annealing and gene conversion do not depend on Ku, and are only slightly impaired by the absence of Mre11–Rad50–p95.)

between these gatekeepers of double-strand break repair is especially pertinent to attempts at gene replacement in mammalian cells—in this case, most of the injected DNA ends up stuck together end to end and inserted in nonhomologous locations. So an understanding of how DNA damage is channelled towards one repair process or another is a key step in future attempts to manipulate the genome. □

James E. Haber is in the Rosenstiel Center and the Department of Biology, Brandeis University, MS029, Waltham, Massachusetts 02454-9110, USA.

e-mail: haber@hydra.rose.brandeis.edu

1. Van Dyck, E., Stasiak, A. Z., Stasiak, A. & West, S. C. *Nature* **398**, 728–731 (1999).

2. Umez, K., Sugawara, N., Chen, C., Haber, J. E. & Kolodner, R. D. *Genetics* **148**, 989–1005 (1998).
3. Sugawara, N. *et al.* *Nature* **373**, 84–86 (1995).
4. Holmes, A. & Haber, J. E. *Cell* **96**, 415–424 (1999).
5. Nishinaka, T., Shinohara, A., Ito, Y., Yokoyama, S. & Shibata, T. *Proc. Natl Acad. Sci. USA* **95**, 11071–11076 (1998).
6. Malkova, A., Ivanov, E. L. & Haber, J. E. *Proc. Natl Acad. Sci. USA* **93**, 7131–7136 (1996).
7. Rattray, A. J. & Symington, L. S. *Genetics* **138**, 587–595 (1994).
8. Rijkers, T. *et al.* *Mol. Cell. Biol.* **18**, 6423–6429 (1998).
9. Mortensen, U. H., Bendixen, C., Sunjevaric, I. & Rothstein, R. *Proc. Natl Acad. Sci. USA* **93**, 10729–10734 (1996).
10. Reddy, G., Golub, E. I. & Radding, C. M. *Mutat. Res.* **377**, 53–59 (1997).
11. Milne, G. T. & Weaver, D. T. *Genes Dev.* **7**, 1755–1765 (1993).
12. Sung, P. *J. Biol. Chem.* **272**, 28194–28197 (1997).
13. Shinohara, A., Shinohara, M., Ohta, T., Matsuda, S. & Ogawa, T. *Genes Cells* **3**, 145–156 (1998).
14. Gasior, S. L., Wong, A. K., Kora, Y., Shinohara, A. & Bishop, D. K. *Genes Dev.* **12**, 2008–2021 (1998).
15. Critchlow, S. E. & Jackson, S. P. *Trends Biochem. Sci.* **23**, 394–398 (1998).

Astronomy

Fine-structure variable?

Antoinette Songaila and Lennox L. Cowie

How constant are the ‘constants’ of nature? Ever since Dirac¹ speculated that temporal changes in the values of various dimensionless physical constants might have occurred over the lifetime of the Universe, vigorous but fruitless attempts have been made to detect such variability. The fine-structure constant, $\alpha = 2\pi e^2/hc$, which characterizes the strength of the electromagnetic attraction between photons and electrons, has attracted particularly aggressive attention. Now an Australian–British group is claiming in *Physical Review Letters*² that it might have detected evidence for a varying α , at a level of about one part in 10^5 , over roughly half the lifetime of the Universe. The authors investigated possible time variation in α using cosmological observations of the absorption spectra of distant (high-redshift) quasars. The result must still be considered extremely preliminary, because of the very real possibility of systematic errors in this type of measurement. But given its importance if correct, it can be expected to provoke considerable scrutiny.

The modern expectation of variation in the fine-structure constant arises from theories such as those unifying gravity with the other forces of nature. The best terrestrial limit on the time variation of α outside the laboratory is based on examination of the decay products of the Oklo phenomenon—an ancient natural fission reactor discovered in 1972 in the Oklo uranium mine in Gabon, West Africa. Analysis³ of the 1.8-billion-year-old decay products gives a range for a fractional change in the fine-structure constant ($\Delta\alpha/\alpha$) of between 0.9×10^{-7} and 1.2×10^{-7} over this period, which would scale linearly to a limit of about one part in 10^6 over the lifetime of the Universe. But there is no reason to expect the time dependence to be linear (an oscillating α is even allowed in

some schemes) and longer time baselines afforded by our ability to measure α by observing the atomic and molecular wavelengths of cosmologically distant objects offer a powerful alternative route.

Wavelength spectra of cosmologically distant quasars provide a natural laboratory for investigating changes in α . Dark, narrow

lines in quasar spectra are produced by absorption of radiation in intervening clouds of gas, many of which are enriched with heavy elements. The wavelength separation between two lines produced by absorption from alkaline atoms and ions (an alkaline doublet) is proportional to α^2 , so any small variation in this separation will be roughly proportional to α , to a first approximation. Because quasar spectra contain doublet absorption lines at a number of redshifts—and so at different times in the history of the Universe—it is possible to check for time variation in α simply by looking for changes in the doublet separation of alkaline-type ions with one outer electron (such as C^{3+} or Si^{3+}), as a function of redshift. Although this sounds straightforward, any change in α will be very small, and so the accuracy of measurement needs to be high.

Not surprisingly, this method has a long history of null measurements stretching back to its first application by Savedoff⁴ in 1956. We obtained the most accurate estimate⁵ on the fractional change $\Delta\alpha/\alpha$, prior to the present paper, of $3.5(\pm 5.5) \times 10^{-6}$. This limit was reached by comparing the absorption lines of neutral hydrogen and carbon in a cloud of intergalactic gas at a redshift (z) of 1.8 lying nearly two-thirds of the way across the Universe. The relative wavelengths of the lines would have been shifted by any change

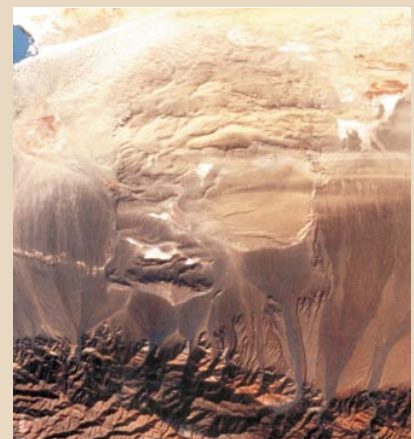
Earth science

A big slip

Examples of past landslides that are known to have been triggered by earthquakes are rarities. But Hervé Philip and Jean-François Ritz believe they have identified one, and it was huge.

From satellite and aerial images of the Gobi-Altay mountain range in Mongolia, and data from ground surveys, Philip and Ritz conclude that the structure shown here, running from the ridge at the bottom of the picture downslope into a valley to the north, is a ‘palaeolandslide’—one of uncertain age, which was triggered by ground-shaking and involved the slippage of 50 km^3 of earth. The authors calculate that some 100 billion tons of material moved which, put another way, is enough to cover The Netherlands waist deep in mud and rock.

The account appears in *Geology* (27, 211–214; 1999), and the authors surmise that there is an earthquake connection for various reasons, most notably the landslide’s occurrence close to the currently active Bogd fault system. The fault ruptured most recently in 1957, in an earthquake of magnitude 8.3. But the most curious aspect of the story is that the average slope from ridge to valley in the landslide area is a very gentle 3° . That



means that, once the ground shaking had stopped, there must have been some unusual conditions for such a mass of material to continue in motion for several kilometres; this picture shows an extent of ground about 34 km from bottom to top, 18 km of which was affected by the slump.

Philip and Ritz suggest that part of the explanation is that there was little or no frictional or cohesive strength at the landslide’s shear plane, a condition that might have pertained when the water table was much higher than it is now. **Tim Lincoln**

CENTRE NATIONAL D'ETUDES SPATIALES 1996/SPOT IMAGE

in the fine-structure constant, but no such change was seen. This is not that surprising: theories allow enormous latitude in the actual variation of α , and a detection at the current level of measurement sensitivity will always be unexpected.

Webb *et al.*² introduce a new variant of this technique that compares the absorption wavelengths of magnesium and iron atoms in the same absorbing cloud, which they demonstrate in a companion paper⁶ to be an order of magnitude more sensitive than the alkaline-doublet method. They observe a number of intergalactic clouds at redshifts from 0.5 to 1.6, seen in absorption against a background of quasars. For the entire sample they find $\Delta\alpha/\alpha = -1.1(\pm 0.4) \times 10^{-5}$, consistent with a null result (to within three standard deviations), but crucially most of the signal comes from a small subset of quasars near a redshift of 1. Restricting the sample to $z > 1$ systems gives a significant change of $\Delta\alpha/\alpha = -1.9(\pm 0.5) \times 10^{-5}$. Such a result would be consistent with little time variation in the past 1.8 billion years since the Oklo event, but larger variation at earlier times when $z > 1$, although it then becomes hard to understand the previous null result⁵ at a redshift of 1.8. Webb *et al.*² also seem unimpressed by the redshift dependence they have apparently uncovered, and interpret their results as "stringent upper limits on any possible time variation rather than a positive detection of change", in part because "a genuine physical effect confined to one specific epoch ... does not seem at present to be well motivated by theoretical expectations".

Of course, the real problem (as Webb *et al.* discuss at length) is the subtle and not-so-subtle systematic errors that, alas, in this type of measurement are potentially legion. The authors mention most of the likely culprits: the wavelength calibration, uncertainties in the laboratory wavelengths of iron and magnesium, and the effect of unresolved velocity substructure in the absorption lines. The redshift dependence of the effect makes the last by far the most likely: it is not hard to imagine that some of their many lines of sight could have hit an absorbing cloud or two in which their assumptions about the relative distribution and kinematics of light and heavy ions were not correct.

Another potential source of systematic error is the use of both very strong and very weak absorption lines of singly ionized iron: any small perturbation would affect the latter much more than the former and could skew the results. Still, for now, and until we can really probe the details of this investigation, subtle effects in the structure of the absorption lines of the two atoms or slight systematic problems in the wavelength calibration remain the most likely reasons. Because of these problems, it is indeed probably best to regard this measurement as an upper limit, albeit a provocative one.

This new technique ultimately could and should be pushed to higher redshift by using shorter-wavelength lines. And the only thing currently limiting the alkaline-doublet method⁵ is the laboratory measurement of the doublet separation of triply ionized silicon lines. The final irony is that both techniques are limited in sensitivity or scope by poorly defined laboratory wavelengths. Webb *et al.* conclude their paper with a plea for more laboratory work to be done, with which we can only concur. □

Antoinette Songaila and Lennox L. Cowie are at the Institute for Astronomy, University of Hawaii, 2680 Woodlawn Drive, Honolulu, Hawaii 96822, USA. e-mails: acowie@uhifa.hawaii.edu
cowie@uhifa.hawaii.edu

1. Dirac, P. A. M. *Nature* **139**, 323 (1937).
2. Webb, J. K., Flambaum, V. V., Churchill, C. W., Drinkwater, M. J. & Barrow, J. D. *Phys. Rev. Lett.* **82**, 884–887 (1999).
3. Damour, T. & Dyson, F. *Nucl. Phys. B* **80**, 37–54 (1996).
4. Savedoff, M. P. *Nature* **178**, 688–689 (1956).
5. Cowie, L. L. & Songaila, A. *Astrophys. J.* **453**, 596–598 (1995).
6. Dzuba, V. A., Flambaum, V. V. & Webb, J. K. *Phys. Rev. Lett.* **82**, 888–891 (1999).

Immunogenetics

Defence by diversity

Adrian V. S. Hill

Twenty five years ago Zinkernagel and Doherty¹ discovered that, for virally infected cells to be destroyed, viral antigens must be presented to killer T cells by class I molecules of the major histocompatibility complex (MHC). This finding suggested how the remarkable diversity of the MHC class I might be selected and maintained in a population. First, certain MHC class I variants — in man, called the human leukocyte antigen (HLA) — A, -B and -C types — might present antigens from a virus more effectively than other variants, generating a stronger, more protective, T-cell response². And second, a variety of HLA types could be selectively maintained by heterozygote advantage³. What this means is that heterozygotes (people who have two, rather than one, types of antigen-presenting molecule per HLA class I locus) could present a wider range of pathogen antigens, and again generate stronger immune responses (Fig. 1). Large-scale studies of HLA class I antigens, reported in *Proceedings of the National*

*Academy of Sciences*⁴ and *Science*⁵, now support the operation of these mechanisms in human viral diseases.

Jeffery *et al.*⁴ report that the common MHC class I type, HLA-A*02, is associated with a greater than two-fold reduction in the risk of disease induced by human T-lymphotropic virus-1 (HTLV-1). In a two-stage study, the frequency of HLA-A*02 was consistently lower in Japanese patients with disease than in infected people without disease. Each of the main HLA-A*02 subtypes present appeared protective, and the authors detected high levels of virus-specific T cells. Among the healthy carriers of HTLV-1, viral load was much less in those with HLA-A*02 — consistent with a protective role for these T cells. The HLA-A2 lineage is unusual in that it is unique to humans, and this study identifies one pathogen that may have contributed to its selection⁶. Carrington *et al.*⁵ report that Americans who are heterozygous for HLA-A, -B or -C antigens are protected against rapid progression to AIDS after

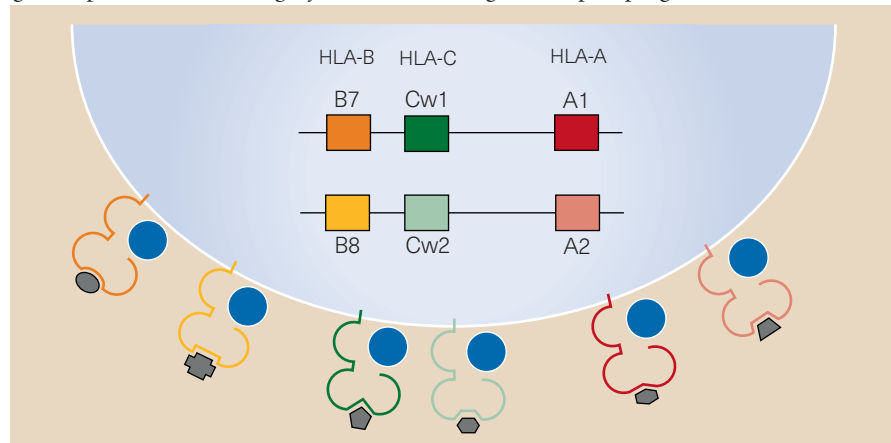


Figure 1 A greater diversity of human leukocyte antigens (HLAs) may protect against some infectious diseases. Jeffery *et al.*⁴ and Carrington *et al.*⁵ have shown that people with certain HLAs are less likely to develop disease after infection with HTLV-1 and HIV-1, respectively, than people without. People with two (rather than one) types of antigen-presenting molecule at each HLA class I gene locus, HLA-A, HLA-B and HLA-C, express six different peptide-presenting molecules on the cell surface. So, these people may generate a stronger immune response against intracellular viral pathogens than those with only one type of antigen-presenting molecule per locus. Virally derived peptides are shown (grey) in the antigen-presenting groove of the HLA class I molecules.

infection with the human immunodeficiency virus-1 (HIV-1). These authors observed an increased risk of rapid disease progression with a reduction in the number of HLA-A, -B and -C types.

These studies support a pathogen-driven theory for variation (or 'polymorphism') of the MHC. Other mechanisms for the maintenance of polymorphism are also gaining credence. Rodents use MHC-determined odour preferences to select mates⁷, and preliminary behavioural studies involving sweaty T-shirts indicate that some human students may use comparable cues⁸. Also, HLA homozygosity (where people have only one HLA type per locus) might impair fetal survival through an unknown mechanism⁹.

Association studies of infectious diseases have often been too small to assess typically low-frequency HLA variants — a problem that is compounded by the extensive diversity at these loci. The more convincing infectious disease associations have usually been with HLA class II antigens¹⁰, such as HLA-DR2, which increases susceptibility to leprosy and tuberculosis in Asia. Jeffery *et al.*⁴ confirm a previous association between the class II antigen HLA-DR1 and disease susceptibility in HTLV-1 infection, and speculate that this may relate to a pathogenic effect mediated by CD4-positive T cells. There is also strong evidence that HLA-DR13 and HLA-DR11 are associated with clearance of infection by the hepatitis B and C viruses, respectively¹⁰. The first evidence for a protective effect of HLA heterozygosity was also found in persistent hepatitis B virus infection but, again, only for HLA class II antigens¹¹.

Why have HLA class I associations been more difficult to identify? Peptide epitopes (the sequences that T cells focus on) tend to bind to a greater variety of class II than class I molecules, suggesting that clearer class I associations with disease should be found. Technical issues may be part of the answer — molecular typing for the many low-frequency class I variants was introduced relatively recently. Indeed, Jeffery and colleagues detected the HLA-A*02 protective association with HTLV-1 owing to the unusually high frequency of this allele. The protective effect of HLA-B*53 against severe malaria¹² was discovered for the same reason. A large variety of alleles also minimizes the number of homozygotes, reducing the power of most studies to evaluate heterozygote advantage.

Most HLA associations with infectious disease probably represent small effects, yet they are important in evolutionary terms. For commonly fatal infectious diseases, modest odds ratios are enough to select and maintain polymorphism. Genetic variation in the infectious pathogen may also be significant. There is substantial sequence variation in HIV, and it may change its immunodomi-

nant epitopes repeatedly during an infection, as viral mutants are selected by immune pressure¹³. Such viral polymorphism may explain the differing results in many HLA studies of HIV and AIDS. In human papillomavirus and malarial infections, HLA associations with molecularly defined strains of pathogen have been identified^{14,15}.

Heterozygote advantage has now been documented for HLA in only two viral diseases^{5,11}, and it was absent in a large malaria study¹². Perhaps it is undermined by the tendency of cellular immune responses to focus on just one or a few major epitopes. The switching of such epitopes in HIV infection¹³, and the tiny size of the hepatitis B virus genome (which leads to a paucity of epitopes), may explain why heterozygote advantage has been observed in these infections. But, in general, other forms of balancing selection might be more important¹⁶. One is frequency-dependent selection, where rare alleles have a selective advantage — possibly because they arose more recently, giving pathogens less time to adapt to them. Another is fluctuation in selection pressures (produced, for example, by intermittent epidemics), which may also maintain polymorphism. Finally, in the presence of numerous infectious diseases, different alleles, protecting against individual pathogens, could make heterozygosity advantageous overall.

The elegant studies of Jeffery *et al.*⁴ and Carrington *et al.*⁵ document the evolutionary selection pressures that viruses can exert on HLA types. They also support the still controversial view that killer T cells are mainly protective, rather than pathogenic, in these diseases. The challenge now is to design vaccines that will allow the host to strike back — and exert new selection pressures for viral diversification. □

Adrian V. S. Hill is at the Wellcome Trust Centre for Human Genetics and the Institute of Molecular Medicine, University of Oxford, Oxford OX3 9DS, UK.

e-mail: adrian.hill@imm.ox.ac.uk

1. Zinkernagel, R. M. & Doherty, P. C. *Nature* **248**, 701–702 (1974).
2. Doherty, P. C. & Zinkernagel, R. M. *Lancet* **1**, 1406–1409 (1975).
3. Doherty, P. C. & Zinkernagel, R. M. *Nature* **256**, 50–52 (1975).
4. Jeffery, K. J. M. *et al.* *Proc. Natl Acad. Sci. USA* **96**, 3348–3353 (1999).
5. Carrington, M. *et al.* *Science* **283**, 1748–1752 (1999).
6. Browning, M. & Krausa, P. *Immunol. Today* **17**, 165–170 (1996).
7. Penn, D. & Potts, W. *Adv. Immunol.* **69**, 411–436 (1998).
8. Wedekind, C. & Furi, S. *Proc. R. Soc. Lond. B* **264**, 1471–1479 (1997).
9. Ober, C., Hyslop, T., Elias, S., Weitkamp, L. R. & Hauck, W. W. *Hum. Reprod.* **13**, 33–38 (1998).
10. Hill, A. V. S. *Annu. Rev. Immunol.* **16**, 593–617 (1998).
11. Thursz, M. R., Thomas, H. C., Greenwood, B. M. & Hill, A. V. *Nature Genet.* **17**, 11–12 (1997).
12. Hill, A. V. S. *et al.* *Nature* **352**, 595–600 (1991).
13. Nowak, M. A. *et al.* *Nature* **375**, 606–611 (1995).
14. Apple, R. J. *et al.* *Nature Genet.* **6**, 157–162 (1994).
15. Gilbert, S. C. *et al.* *Science* **279**, 1173–1177 (1998).
16. Apanius, V., Penn, D., Slev, P. R., Ruff, L. R. & Potts, W. K. *Crit. Rev. Immunol.* **17**, 179–224 (1997).

Daedalus

Spiritual matters

The connection between matter and spirit has been debated for millennia. The central mystery is that certain material objects (human beings) contrive to be conscious and to possess a spiritual dimension. This implies that matter itself has some rudimentary spiritual character.

Daedalus reckons that the spiritual world occupies the same space as the material one, but at most points is very weakly coupled to it. Nonetheless, over cosmological space and time, this weak coupling must have brought the two realms into thermodynamic equilibrium. So the spirit world will have acquired the average temperature of the physical universe — 3 K, the temperature of the microwave background. (This explains why a ghost, an invasion of the material world by the spirit one, tends to cool the room.)

So Daedalus is looking for the weak thermodynamic coupling between the two worlds. A warm material object, even in a perfect vacuum and surrounded by a perfect reflector, should slowly cool down by thermal leakage into the spiritual world. The experiment will test modern thermometry and high-vacuum techniques to their limits; but success would open a whole new field of discovery.

Daedalus is already planning his exploration of it. He hopes to discover whether holy relics and ritual objects, imbued with spiritual significance, cool faster than more mundane ones. He expects that biological materials, especially the neurotransmitters and proteins of brain chemistry, will cool faster still. They are part of the secret of human consciousness, and should be quite tightly coupled to the spiritual world. This coupling might take the form of strong spectroscopic emission at the 3 K black-body peak in the millimetre-wave region, or it might be far more subtle. But detailed experiments, aided by such insights, should reveal the spiritual capacity of a wide range of objects and substances.

Perhaps the most interesting objects for testing will be semiconductors, either as bulk solids or in the form of integrated circuits and microprocessors. If they cool no faster than common minerals, this will demonstrate their lack of a spiritual dimension — in which case no computer, however powerful, could be conscious. But if they turn out to cool as fast as the neurotransmitters, then conscious technological monsters like Hal and Deep Thought should indeed be possible.

David Jones

Obituary

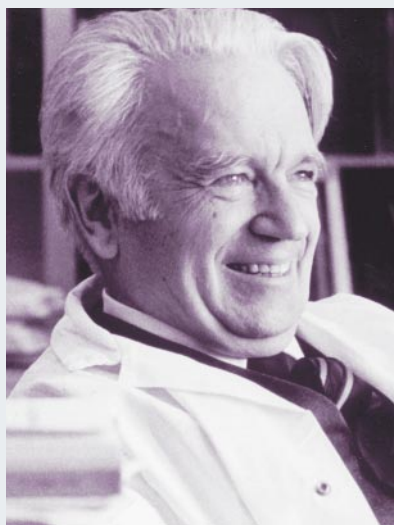
Gerhard Herzberg (1904–99)

Patriarch of modern molecular spectroscopy

Gerhard Herzberg, who died on 3 March, was one of the world's foremost molecular spectroscopists. The myriad wavelengths of light emitted or absorbed by molecules offer a vast lode of knowledge about the arrangement of atoms within molecules, the properties of the chemical bonds linking atoms, and the electronic interactions governing those bonds. Moreover, insight into how bonds can be made or broken in chemical reactions is crucial for understanding phenomena in physics and astrophysics, chemistry, materials science, biology and medicine. Developing the means to observe and decipher molecular spectral patterns has been a mighty task, in which Herzberg had a leading role for 60 years.

Herzberg was born on Christmas day in 1904 in Hamburg, Germany. When he was a young student, inspiring teachers encouraged his interest in science. However, after the director of the Hamburg Observatory dissuaded him from pursuing advanced study in astronomy, he entered the Technische Hochschule in Darmstadt, where he completed a doctorate in engineering physics in 1928. The period of his doctoral work coincided with the advent of quantum mechanics, the essential theoretical foundation for molecular spectroscopy. Herzberg liked to say "this fortunate accident made it possible for me to learn the subject while it was being developed".

At Göttingen, then one of the prime centres for quantum mechanics, Herzberg worked both in the experimental institute headed by James Franck and the theoretical institute led by Max Born. During that time, Herzberg made his first measurements of electronic transitions in polyatomic molecules. He also did important theoretical work. In particular, in collaboration with Walter Heitler, Herzberg made use of a theorem, proven by Eugene Wigner a few months before, to show that intensity variations seen in rotational lines of the spectrum of the diatomic nitrogen molecule were in conflict with the prevailing view of nuclear structure. This was a striking result, not cleared up until three years later, when James Chadwick discovered the neutron. Herzberg also began work on molecular orbital theory as a systematic approach to describing the electronic configurations of molecules and their spectral transitions.



At the invitation of J. E. Lennard-Jones, who was also working on molecular orbital theory, Herzberg continued his studies of spectra at the University of Bristol, and diligently practised his English. He submitted his work on molecular orbitals as a habilitation thesis to Darmstadt and in 1930 was appointed *Privatdozent* there. This meant he did not receive a salary, but was able to earn a modest income by supervising laboratories and research students. He pursued research vigorously, and collaborated with Edward Teller in working out selection rules for the vibrational structure of electronic transitions of polyatomic molecules. When deuterium was discovered by Harold Urey in 1932, Herzberg recognized how this isotope of hydrogen might benefit studies of molecular structure. He demonstrated this by synthesizing several deuterium-labelled compounds and analysing the spectral shifts induced by replacing hydrogen with deuterium.

After Hitler came to power in 1933, anyone with a Jewish wife was not allowed to teach at a German university, and Herzberg looked for a position in another country. In 1935, the Herzbergs relocated to the University of Saskatchewan in Saskatoon, Canada. This haven came about because of the efforts of a young physical chemist, J. W. T. Spinks, who had worked with Herzberg on deuterated molecules. Here Herzberg began some of his most seminal work, including experiments in which he identified molecules in interstellar space by replicating their spectra in the laboratory.

In 1945 Herzberg went to the Yerkes Observatory of the University of Chicago, where he discovered the quadrupole

spectrum of molecular hydrogen, destined to have a major role in identifying hydrogen both in interstellar space and in planetary and stellar atmospheres. In 1948 he returned to Canada to join the National Research Council laboratory (NRC) in Ottawa, in which he served as director of physics for 20 years. During his era, the NRC became the foremost laboratory for molecular spectroscopy in the world, by virtue of his own fruitful work, and the enterprising spirit and creative freedom he fostered in the research staff.

Herzberg was a mentor to legions of scientists by means of his definitive series of books on atomic and molecular spectroscopy, published in five volumes between 1936 and 1979. Extraordinary in scope, yet readable by the neophyte, his books have been translated into several languages. Researchers find them essential guides, with their comprehensive tables of molecular parameters and abundant illustrations of characteristic spectra and diagrams explaining theoretical concepts.

Herzberg's unpretentious, forthright manner, integrity and devotion to science were greatly admired by colleagues and students. His voice was a powerful baritone, enhanced by singing lessons with which he rewarded himself after completing his first book. Also relished were his ready humour and storytelling. A revealing tale stemmed from a visit to the Soviet Union in 1959. His books had been reprinted for Soviet scientists, and at his request the royalties were delivered: a shoebox full of roubles that could only be spent within the Soviet Union. Cheerfully, he applied a strategy akin to that he had used when leaving Nazi Germany over 25 years before. With the roubles he purchased several round-the-world aeroplane tickets, which he later exchanged for cash.

Herzberg received many honours, including the 1971 Nobel Prize in Chemistry. Particularly cited were his intrepid studies of free radicals, extremely reactive chemical species that are transient intermediates in many reaction processes. In 1975 the NRC Herzberg Institute of Astrophysics was created in his honour, and he continued active research until the age of 90. His legacy will long endure, and continue to grow, as methods he developed or elucidated enable new generations of scientists to read intriguing messages encoded in molecular spectra.

Dudley Herschbach

Dudley Herschbach is in the Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, Massachusetts 02138, USA.
e-mail: herschbach@chemistry.harvard.edu

Antibacterial peptide from *H. pylori*

Colonization of the human stomach by the bacterium *Helicobacter pylori* is a predisposing factor for gastrointestinal illnesses such as gastritis and peptic ulcers¹. But most infections are asymptomatic, and it has recently been suggested that *H. pylori* may actually have beneficial effects on infected carriers who are heavily exposed to other gastrointestinal pathogens². We find that *H. pylori* possesses antibacterial activity to which it is itself resistant. We have traced this activity to cecropin-like amino-terminal peptides derived from the ribosomal protein L1 (RpL1).

The antibacterial activity was detected when a crude lysate of *H. pylori* was tested in the growth-inhibition zone assay³. The activity is protease sensitive, suggesting that *H. pylori*, like some other bacteria⁴, could produce one or more antimicrobial peptides⁵. A consensus motif $(-(\text{knd})-\text{f}-\text{f}-\text{k}-(\text{kre})-(\text{il})-\text{e}-(\text{kr})-\text{f}-\text{f}-\text{x}-(\text{hkrn})-(\text{ivt})-(\text{rkqn})-(\text{dn})-$ for 15 N-terminal residues from insect cecropins was used to search the SWISS-PROT database. Only one additional sequence was found: the N terminus of RpL1 from *H. pylori*. A consensus motif for defensins generated no positive hits in the database.

Cecropins are antibacterial peptides which are composed of two amphipathic α -helices joined by a hinge⁶. Unlike several RpL1 proteins from other bacteria, the RpL1 N terminus of *H. pylori* has the ability to form a perfect amphipathic helix. This first helix (residues 2–19) is followed by a second helix (residues 22–38).

Four peptides corresponding to the N-terminal part of *H. pylori* RpL1 were synthesized and tested for antibacterial activity. Two peptides, Hp(2–20) and Hp(22–38), were chosen to correspond to each of the anticipated α -helices. The other two, Hp(2–13) and Hp(2–38), correspond to a truncated first helix and to both helices,

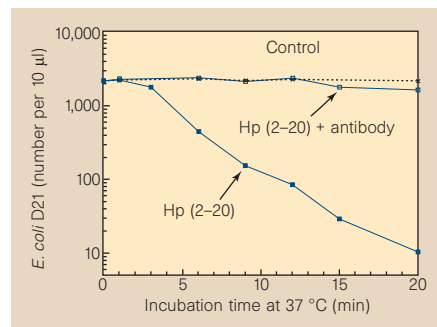


Figure 1 Time curve for the killing of *Escherichia coli* strain D21 by peptide Hp(2–20). Preincubation with affinity-purified antibodies against peptide Hp(2–20) blocked the killing effect. Incubation mixtures contained 2×10^6 bacteria and $0.4 \mu\text{g}$ peptide in $100 \mu\text{l}$ phosphate-buffered saline.

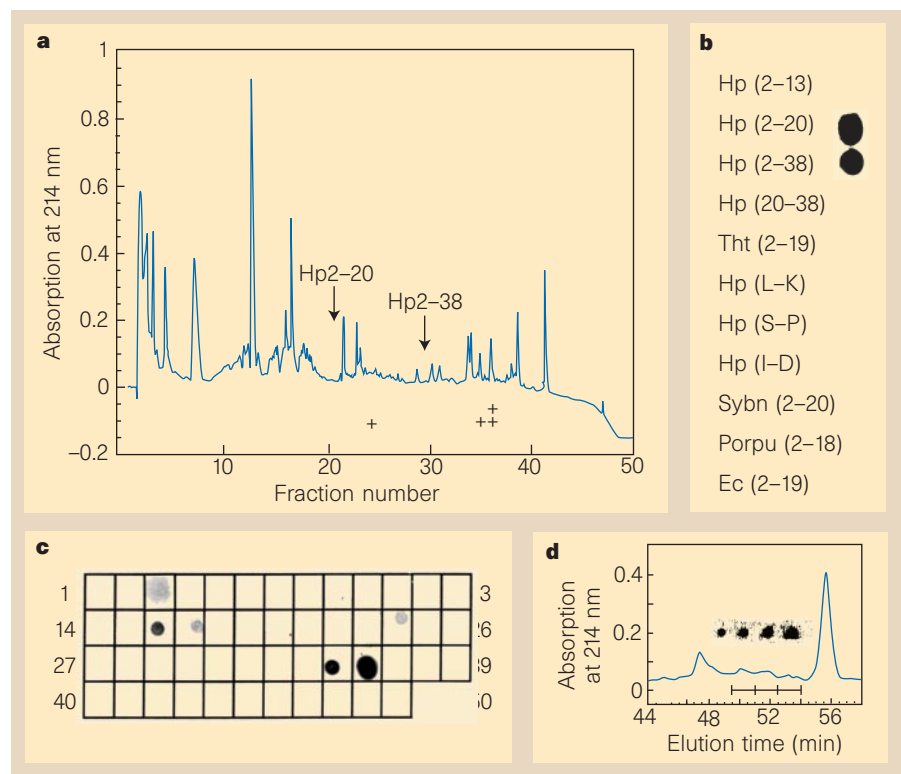


Figure 2 High-performance liquid chromatography separation of antibacterial components in an acetonitrile-extracted *H. pylori* strain MO19 lysate. **a**, Of 50 fractions collected (0.8 ml min^{-1}), concentrated and assayed, antibacterial activity (+) was found in three fractions (main activity in tube 36 (+ +)). Arrows indicate elution times of the synthetic peptides Hp(2–20) and Hp(2–38). **b**, Reactivity with anti-Hp(2–20) antibody using the dot-blot assay gave three strong and three weak signals. **c**, Specificity of antisera illustrated by a dot-blot assay (fraction numbers are shown). **d**, Magnified part of a rechromatogram showing the main antibacterial activity.

respectively. All peptides except Hp(20–38) exerted antibacterial activity against the Gram-negative *Escherichia coli* strain D21 and the Gram-positive *Bacillus megaterium* strain Bm11. The rate of killing of *E. coli* D21 with peptide Hp(2–20) was 99.5% in 20 minutes. Preincubation of the peptide with purified antibodies raised against the peptide blocked this activity (Fig. 1). The same antibody inhibited the main antibacterial activity of the crude *H. pylori* lysate.

H. pylori RpL1 peptides and cecropins A and P1 were inactive against all *H. pylori* strains tested. It has been suggested that cecropin A binds to the diphosphoryl lipid A moiety of lipopolysaccharide⁷. It may be the lower amount of phosphates in *H. pylori* lipid A⁸ that makes it resistant to cecropin.

We extracted the antibacterial agent from the *H. pylori* lysate by using 0.1% trifluoroacetic acid in 60% acetonitrile (final concentration) and fractionated it using reversed-phase high-performance liquid chromatography. Fractions were assayed for antibacterial activity and for reactivity with an affinity-purified anti-Hp(2–20) antibody. Both the main antibacterial activity and the strongest dot-blot signal were

obtained in fraction 36 (Fig. 2a). Replacing amino acids in Hp(2–20) abolished all signals (Fig. 2b), indicating that the anti-Hp(2–20) antibody is highly specific. The elution times for the synthetic peptides Hp(2–20) and Hp(2–38) are marked by arrows in Fig. 2a. From these results, we conclude that the main antibacterial agent (fraction 36) is mediated by an N-terminal fragment(s) of RpL1 that is probably larger than the synthetic peptide Hp(2–38). The weak dot-blot signals obtained in fractions 3, 16 and 17 may originate from shorter RpL1 peptides at concentrations too low to have detectable antibacterial activity.

We pooled the active fractions for rechromatography using a flatter gradient to improve the resolution. A magnified part of such a chromatogram with dot-blot signals detected in four fractions is shown in Fig. 2d. Fractions indicated by a horizontal bar were analysed by mass spectrometry, but no clear interpretation of the mass data could be obtained. Our data are compatible with the presence of several N-terminal fragments of RpL1 in the lysate.

H. pylori has been demonstrated to undergo 'altruistic lysis' *in vivo*⁹. We suggest

that bacterial lysis in the stomach could release L1-derived cecropin-like peptides that are active against faster-growing microorganisms found there. Our data are also consistent with the idea that cecropins have evolved from an early *rpl1* gene in a prokaryote that passed from being an intracellular parasite to a symbiont, ending up as an organelle. When the *rpl1* gene moved from the organelle to the host nucleus, a duplicated sequence could have begun to evolve towards a specialized antimicrobial peptide.

**Katrin Pütsep, Carl-Ivar Brändén,
Hans G. Boman, Staffan Normark**

Microbiology and Tumour Biology Centre,
Karolinska Institutet, Box 280,
SE-171 77 Stockholm, Sweden
e-mail: hans.g.boman@swipnet.se

- Blaser, M. *Br. Med. J.* **316**, 1507–1510 (1998).
- Mattsson, A., Lönnroth, H., Quiding-Järbrink, M. & Svennerholm, A.-M. *J. Clin. Invest.* **102**, 51–56 (1998).
- Hultmark, D. *et al. EMBO J.* **2**, 571–576 (1983).
- Boman, H. G. *Scand. J. Immunol.* **48**, 15–25 (1998).
- Sahl, H. G. & Bierbaum, G. *Annu. Rev. Microbiol.* **52**, 41–79 (1998).
- Holal, T. A. *et al. Biochemistry* **27**, 7620–7629 (1988).
- DeLuca, A. J., Jacks, T. J. & Brogden, K. A. *Mol. Cell Biochem.* **151**, 141–148 (1995).
- Suda, Y. *et al. J. Biochem.* **121**, 1129–1133 (1997).
- Phadnis, S. H. *et al. Infect. Immun.* **64**, 905–912 (1995).

Cleaner fish really do clean

The cleaning of client fish by cleaner fish is one of the most highly developed interspecific communication systems known. But even though it is a seemingly obvious mutualism^{1,2}, several quantitative studies^{3–5} have failed to show any benefit for the clients, leading to the hypothesis that cleaner fish are 'behavioural parasites' that exploit the sensory system of the clients⁶ to obtain food, rather than to increase the client's fitness. The cleaner fish *Labroides dimidiatus* eats parasitic gnathiid isopods, which decline in number on the client fish *Hemigymnus melapterus* daily between dawn and sunset^{7,8}. I find that the cleaner fish reduces parasite abundance, resulting in a 4.5-fold difference within 12 hours, supporting the hypothesis that cleaning behaviour is mutualistic.

Various models have been proposed to explain cooperative interactions among unrelated individuals⁹. The Iterated Prisoner's Dilemma is the generally accepted model¹⁰, although it has been criticized². The behaviour of cleaner fish has been used to highlight the limitations of repeated games, and may be useful for developing alternative models, because if the client cheats by eating the cleaner, the game is over, so it is not a repeated game². Understanding the benefits of cleaning (clients

have fewer parasites) is essential for testing such models.

I compared the number of parasitic gnathiids on caged *H. melapterus* on reefs with and without cleaner fish. Measurements were taken after 12 days at sunset to examine the long-term effect of cleaner fish, and after 12 and 24 hours, at dawn and sunset, respectively, to determine whether the observed decline in gnathiid abundance⁸ over the day was caused by cleaner fish.

Six cages, each containing three *H. melapterus*, were placed on each of five small coral reefs adjacent to Lizard Island, Great Barrier Reef (reefs 7, 8, 14, 15 and 16 on map in ref. 4), of which three reefs had all cleaner fish, *L. dimidiatus*, removed. Cages were put out at dawn on a different reef, randomly selected, each day, and prawns were fed to fish daily. Fish on reefs 7 and 8 were cleaned regularly after cages had been on the reefs for 9 and 8 days, respectively. In the first experiment, after 12 days, the fish were recovered at sunset and gnathiids counted. The effect of cleaner fish on gnathiid abundance was tested (Fig. 1).

In the 24-hour experiment, the above fish were returned to the holding tanks until the following sunset, when they were placed on the same reef; fish from the first experiment that were missing or dead were substituted with other fish from the laboratory. Half the cages on each reef were recovered the following dawn, and the rest were recovered the next sunset, and gnathiids were counted again. Gnathiid abundance was tested for the effects of cleaner fish and sampling time (Fig. 1). The effect of replacing the missing fish on gnathiid abundance was also tested and was not significant ($F_{1,52} = 1.97$, $P = 0.166$). Variability in gnathiid abundance between reefs from dawn to sunset, within a treatment, was found to change significantly (analysis of covariance, $F_{3,19} = 4.10$, $P = 0.021$), probably because there were high gnathiid loads at both dawn and sunset at reef 16, so I explored the effect of time on gnathiid abundance by reanalysing each time separately.

Cleaner fish had a clear effect on the abundance of parasites. After 12 days and at sunset, fish on reefs without cleaner fish had on average 3.8 times more gnathiids than fish on reefs with cleaners (Fig. 1a). In the 24-hour experiment, gnathiid abundance did not differ between treatments at dawn (Fig. 1b); in contrast, gnathiids on reefs without cleaner fish and sampled at sunset had 4.5 times more gnathiids than fish on reefs with cleaner fish (Fig. 1c).

Cleaner fish eat 1,200 parasites per day (mostly gnathiids) and feed only during the day⁷, whereas gnathiids infest fish during both day and night⁸. The rapid reduction in gnathiid abundance between dawn and sunset indicates that the daily decline in

gnathiids is probably due to cleaner fish. Similar quantitative studies, in which the cleaner fish were removed for periods from one month to two years, found no effect of cleaners on parasite abundance^{3–5}. This may be due to the different client and cleaner fish species used, the presence of other unidentified cleaners, movement between reefs by clients, and spatial or temporal variation in parasite loads.

My results show that cleaner fish have an effect both within 12 days and within 12 hours. Cleaner fish are known to benefit from cleaning⁷, but my data show that they greatly reduce the abundance of gnathiids, which can be deleterious to fish¹¹. This is consistent with the hypothesis that cleaning behaviour is mutually beneficial to both participants, and paves the way to using

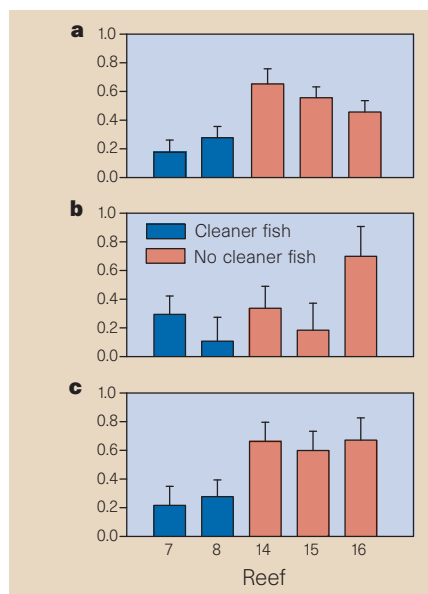


Figure 1 Gnathiids on caged fish on reefs with and without cleaner fish. Data are least-square means and standard errors for fish sampled at the following times: **a**, at sunset after 12 days ($F_{13} = 17.64$, $P = 0.0246$); **b**, at dawn after 12 hours ($F_{19} = 1.80$, $P = 0.213$; reef (treatment), $F_{3,19} = 2.15$, $P = 0.164$); **c**, at sunset after 24 hours ($F_{13} = 11.56$, $P = 0.042$; reef (treatment), $F_{3,13} = 0.11$, $P = 0.950$). Variation in gnathiid abundance in **a** was tested with a nested analysis of covariance (ANCOVA) with presence of cleaners as the main effect and reefs nested within treatments; the error term testing for an effect of treatment was the type III mean square for reef (treatment). In the 24-hour experiment, a multifactor nested ANCOVA was used; the effect of using replaced fish was tested by including it as a factor; the error term used to test the main effects and interactions was the type III means square for cage (treatment $\times$ time $\times$ reef). Because there was a significant interaction of time $\times$ reef (treatment), separate ANCOVAs were used for each time with fish size as the covariate. Gnathiid abundance and fish size were $\log_{10}(x+1)$ and $\log_{10}(x)$ transformed, respectively. Least-square means were calculated to show gnathiid abundance while accounting for fish size.

cleaning behaviour to test models of non-kin cooperation.

Alexandra S. Grutter

Department of Microbiology and Parasitology,
University of Queensland,
Brisbane, Queensland 4072, Australia
e-mail: a.grutter@mailbox.uq.edu.au

1. Trivers, R. L. *Q. Rev. Biol.* **46**, 35–57 (1971).
2. Hammerstein, P. & Hoekstra, R. F. *Nature* **376**, 121–122 (1995).
3. Losey, G. S. *Symbiosis* **4**, 229–258 (1987).
4. Grutter, A. S. *J. Exp. Mar. Biol. Ecol.* **196**, 285–298 (1996).
5. Grutter, A. S. *Oecologia* **111**, 137–143 (1997).
6. Losey, G. S. *Anim. Behav.* **27**, 669–685 (1979).
7. Grutter, A. S. *Mar. Biol. Prog. Ser.* **130**, 61–70 (1996).
8. Grutter, A. S. *Mar. Biol.* (in the press).
9. Clements, K. C. & Stephens, D. W. *Anim. Behav.* **50**, 527–535 (1995).
10. Dugatkin, L. A. *Cooperation Among Animals* (Oxford Univ. Press, 1997).
11. Paperna, I. & Por, F. D. *Rapp. Comm. Int. Mer. Médit.* **24**, 195–197 (1977).

Perception of changes in loudness

Neuhoff¹ reported that “rising level tones... change (in loudness) more than falling level tones despite having the same actual change in level... indicating that direction of change is an important (and previously unaddressed) factor in the perception of dynamic loudness change”, and speculated that: “In a natural environment this overestimation could provide a selective advantage, because rising intensity can signal movement of the source towards an organism.” Leaving aside the question of why it may not be as important for survival to detect the movement of a sound source away from an organism, we dispute the assertion that there is no prior evidence about the influence of direction of change on the degree of change in perceived loudness. This evidence does exist and shows, in contrast to the result reported by Neuhoff¹, that declining signal intensity covers a greater range of loudness than does rising signal intensity.

Twelve years earlier it was reported² that tones continuously decreasing in intensity from moderate to very low levels follow a course of accelerated “softening” until the end of the downsweep, when the loudness is very much less than it would be if that final intensity had been presented alone. The underlying mechanism for this phenomenon, which was subsequently named “decrement”³, is unknown, but arguments have been made for both central and peripheral factors. For example, it was later shown⁴ that at least some part of the effect occurs at the receptor site, as a test in the contralateral ear at the end of the downsweep showed little or no decrement.

In these earlier studies, the loudness of sounds sweeping up, instead of down, over the same ranges of intensity yielded some

evidence for loudness enhancement (or “upcruitment”) at the end of the sweep. However, the effect, where it occurs at all, is much smaller than the accelerated softening in downsweeps. It should be noted that the earlier work concerned sweep durations much longer than the 1.8 seconds used by Neuhoff. A more recent study^{5,6} has shown that, although decrement is diminished when duration is as short as one second, a 30-decibel downsweep for 1- and 2.5-second durations covers a bigger range of loudness than the comparable upsweep.

It would therefore be prudent to limit Neuhoff’s conclusions to his own test conditions and method of measuring loudness change: the larger loudness change recorded earlier for downsweeps under a range of conditions and with a variety of measurement techniques contrasts with Neuhoff’s finding based on a single duration, a single intensity range, and a single measurement procedure.

How can this apparent conflict be explained? Perhaps, as Neuhoff argues, direct judgement of “perceived change” taps a fundamentally different process from judgements of loudness obtained at the beginning and end of a sweep. But listeners have been known to make their own decisions about what to attend to in an experimental setting: although Neuhoff cautioned his subjects to avoid making a judgement of overall loudness, they may have done so; the direction of change may have an effect on such judgements that is quite different from its effect on loudness judgements made at discrete stages of a sweep. Close attention should also be paid to procedural differences that may turn out to be decisive.

Neuhoff’s results are of great interest but, in our view, they need to be considered in the context of what is already known about loudness perception for signals that are continuously changing in intensity. It remains to be seen whether his finding is of sufficient generality to warrant speculation about its role in the process of evolutionary selection.

Georges Canévet*, Bertram Scharff†, Robert S. Schlauch‡, Martha Teghtsoonian§, Robert Teghtsoonian§

*Laboratoire de Mécanique et d’Acoustique, CNRS, 13402 Marseille, France

†Auditory Perception Laboratory, Northeastern University, Boston, Massachusetts 02115, USA

‡Department of Communication Disorders, University of Minnesota, Minneapolis, Minnesota 55455, USA

§Department of Psychology, Smith College, Northampton, Massachusetts 01063, USA
e-mail: rtx@smith.edu

1. Neuhoff, J. *Nature* **395**, 123–124 (1998).
2. Canévet, G. *Acustica* **61**, 256–264 (1986).
3. Canévet, G. & Scharf, B. *J. Acoust. Soc. Am.* **88**, 2136–2142 (1990).

4. Schlauch, R. S. *J. Acoust. Soc. Am.* **92**, 758–765 (1992).
5. Teghtsoonian, R., Teghtsoonian, M. & Canévet, G. in *Fechner Day 95* (ed. Possamai, C.-A.) 309–314 (International Society for Psychophysics, Cassis, 1995).
6. Teghtsoonian, R., Teghtsoonian, M. & Canévet, G. *Percept. Psychophys.* (in the press).

Neuhoff replies — Canévet and his colleagues suggest that their findings address dynamic loudness change and are inconsistent with my recent discovery of a bias for rising intensity tones¹. However, the two sets of experiments address fundamentally different questions. Canévet *et al.*’s listeners were asked to make judgements about loudness, whereas my listeners were asked to make judgements about the amount of dynamic change. Essentially, their listeners answered the question “How loud is it now?” by assigning a number to the loudness of a changing intensity sound at various times throughout the stimulus duration. This provided a discrete measure of loudness at various snapshots in time. Listeners in my experiments were specifically asked to ignore the overall loudness of the sounds and to make summary judgements about the amount of loudness change, essentially answering the question “How much did it change in loudness?”.

From static judgements of loudness at discrete points in time, Canévet and his colleagues infer the perceived amount of dynamic loudness change. But they do not measure perceived dynamic change directly. There can be inherent shortcomings in inferring characteristics of dynamic perception by extrapolating from static judgements. Their experiments are important and have implications in areas from physiology to auditory display. However, once the difference between the perception of loudness and the perception of dynamic loudness change is made clear, perhaps the discrepancy between our results is not so unexpected.

Judgements of dynamic change may be mediated by different mechanisms from static judgements of loudness. Because intensity change is the important factor in specifying the arrival time of a source², a direct judgement about the amount of change might be more useful for localizing a source than a series of snapshot judgements of loudness. A bias for rising intensity when judging loudness change is consistent with studies that show that listeners systematically err on the side of safety when estimating the arrival time of a sound source^{3,4}. The bias for rising intensity may therefore provide a selective advantage.

Canévet *et al.* suggest that their findings undermine this claim. However, keeping in mind the differences between loudness and loudness change, their work is not necessarily inconsistent with the evolutionary position. They have shown for tones that

continuously decrease in intensity that the terminal loudness is less than it would be if that final intensity were presented alone. Such decreasing loudness may signal decreasing environmental importance, because it is consistent with the departure of a sound source. So, viewing their findings in environmental terms, the endpoint of a downward-sweeping sound (or departing source) might be less important (and less loud) than the endpoint presented alone, which could signal a new source.

An analogous effect occurs in vision where an unimportant background in a visual scene can appear to be darker than a more important figure, despite the two having equal luminance⁵. More experiments on the perception of both loudness and loudness change are called for, but for the moment the evolutionary position seems able to accommodate both sets of data.

John G. Neuhoff

Department of Psychology, Lafayette College,
Easton, Pennsylvania 18042-1781, USA
e-mail: neuhoffj@lafayette.edu

1. Neuhoff, J. G. *Nature* **395**, 123–124 (1998).
2. Shaw, B. K., McGowan, R. S. & Turvey, M. T. *Ecol. Psychol.* **3**, 253–261 (1991).
3. Schiff, W. & Oldak, R. J. *Exp. Psychol. Hum. Percept. Perform.* **16**, 303–316 (1990).
4. Rosenblum, L. D., Wuestefeld, A. P. & Saldana, H. M. *Perception* **22**, 1467–1482 (1993).
5. Kanizsa, G. *Organization in Vision: Essays on Gestalt Psychology* (Praeger, New York, 1979).

Movement of motor and cargo along cilia

Intraflagellar transport (IFT)¹ is important in the formation and maintenance of many cilia, such as the motile cilia that drive the swimming of cells and embryos², the nodal cilia that generate left–right asymmetry in vertebrate embryos³, and the sensory cilia that detect sensory stimuli in some animals⁴. The heterotrimeric kinesin-II motor protein drives the anterograde transport of macromolecular complexes, called rafts, along microtubule tracks from the base of the cilium to its distal tip⁵, whereas cytoplasmic dynein moves the rafts back in the retrograde direction⁶. We have used fluorescence microscopy to visualize for the first time the intracellular transport of a motor and its cargo *in vivo*. We observed the anterograde movement of green fluorescent protein (GFP)-labelled kinesin-II motors and IFT rafts within sensory cilia on chemosensory neurons in living *Caenorhabditis elegans*.

The anterograde IFT motor, heterotrimeric kinesin-II⁷, consists of two heterodimerized kinesin-related motor subunits and one accessory subunit (KAP)⁸. To observe kinesin-II-driven IFT within

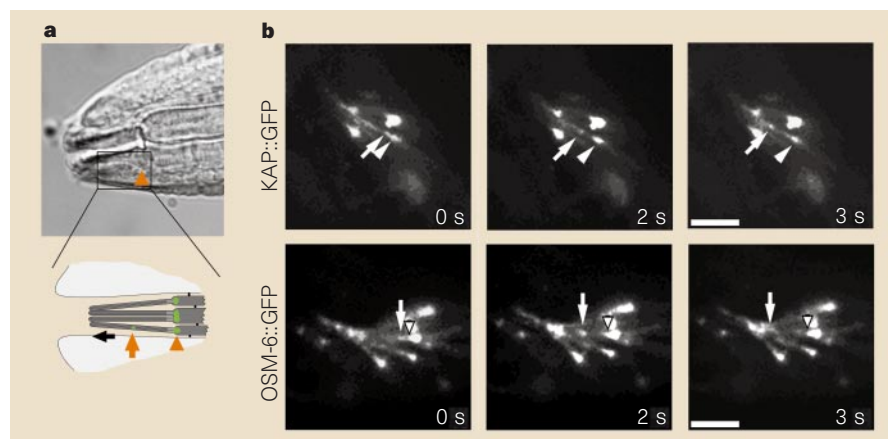


Figure 1 Visualization of intraflagellar transport (IFT). **a**, Schematic diagram of *Caenorhabditis elegans* chemosensory cilia. Top, a differential interference contrast light micrograph of the head of an adult worm orientated to the left. The boxed region shows the position of an amphid channel containing sensory cilia. This region is represented in the diagram (bottom), which shows a close-up of sensory cilia orientated with their distal endings facing left, and contacting the external environment through openings in the cuticle. Green indicates fluorescent kinesin-II or OSM-6, which accumulate at the transition zone at the base of the cilia (orange arrowhead) and move (black arrow) to the distal endings of the cilia. **b**, Fluorescence micrographs of sensory cilia in GFP transgenic worms as represented by the boxed region and the diagram in **a**. Fluorescent kinesin-II motors and fluorescent OSM-6 subunits of IFT rafts accumulate at the base of the transition zones where they appear as large dots (arrowheads). Arrows indicate position of dots of fluorescent kinesin-II and fluorescent OSM-6 as they travel to the distal tip of the sensory cilium (left, as in **a**). A movie of this process can be seen at <http://www.mcb.ucdavis.edu/faculty-labs/scholey/>. Details of the methods are available on request from J. M. S.

chemosensory cilia, we used transgenic lines of *C. elegans* expressing GFP fused to the kinesin-II KAP and to a presumptive cargo molecule, OSM-6, a component of IFT rafts that has an essential role in chemosensory ciliary function^{5,9}.

With a fluorescence microscope, we observed that KAP::GFP and OSM-6::GFP polypeptides accumulate in the region of the transition zone at the base of the sensory cilia. This is consistent with previous immunofluorescence data on IFT motors and raft polypeptides in other systems⁵ (Fig. 1). We observed small fluorescent dots corresponding to the kinesin-II KAP and the OSM-6 cargo emerging from these regions and moving out towards the distal tip of the sensory cilia. Both the motor and its presumptive cargo moved anterogradely at identical rates ($0.65 \pm 0.11 \mu\text{m s}^{-1}$ ($n=50$) for the KAP compared with $0.65 \pm 0.10 \mu\text{m s}^{-1}$ ($n=50$) for OSM-6), which is similar to the velocity of microtubule motility driven by purified heterotrimeric kinesin-II in a motility assay⁷. In contrast, the sensory ciliary transmembrane receptor ODR-10 moved at a faster rate ($1.59 \pm 0.28 \mu\text{m s}^{-1}$ ($n=10$)), confirming that the identical velocities displayed by KAP::GFP and OSM-6::GFP are not an artefact of the recording technique.

This direct viewing of the intracellular transport of a motor and its cargo *in vivo* provides strong support for the hypothesis that heterotrimeric kinesin-II is the motor protein that drives anterograde IFT⁵. In chemosensory neurons of *C. elegans*, it is

likely that kinesin-II-driven IFT delivers structural components of sensory ciliary axonemes and components of the sensory signalling machinery that are concentrated in these cilia.

Genetic studies have identified 25 genes, including *osm-6*, that are essential for ciliary function in this system¹⁰, and the ability to view IFT in organisms carrying mutations in these genes will make it possible to determine which of the corresponding gene products are linked to the kinesin-II transport pathway. In a broader context, our approach should allow the direct observation of motor and cargo molecules participating in IFT in a broad range of cilia and flagella.

Jose T. Orozco, Karen P. Wedaman, Dawn Signor, Heather Brown, Lesilee Rose, Jonathan M. Scholey

Section of Molecular and Cellular Biology,
University of California, Davis,
California 95616, USA
e-mail: jmscholey@ucdavis.edu

1. Rosenbaum, J. L., Cole, D. G. & Diener, D. J. *Cell Biol.* **144**, 385–388 (1999).
2. Morris, R. L. & Scholey, J. M. *J. Cell Biol.* **138**, 1009–1022 (1997).
3. Nonaka, S. *et al.* *Cell* **95**, 829–837 (1998).
4. Perkins, L. A., Hedgecock, E. M., Thompson, J. N. & Culotti, J. G. *Dev. Biol.* **117**, 456–487 (1986).
5. Cole, D. G. *et al.* *J. Cell Biol.* **141**, 993–1008 (1998).
6. Pazour, G. J., Wilkerson, C. G. & Witman, G. B. *J. Cell Biol.* **141**, 979–992 (1998).
7. Cole, D. G. *et al.* *Nature* **366**, 268–270 (1993).
8. Signor, D., Wedaman, K. P., Rose, L. S. & Scholey, J. M. *Mol. Biol. Cell* **10**, 345–360 (1999).
9. Collet, J., Spike, C. A., Lundquist, E. A., Shaw, J. E. & Herman, R. K. *Genetics* **148**, 187–200 (1998).
10. Starich, T. A. *et al.* *Genetics* **139**, 171–188 (1995).

Drugs wipe out a sporting chance

Dying to Win: Doping in Sport and the Development of Anti-Doping Policy

by Barrie Houlihan

Council of Europe Publishing: 1999. 210 pp.

£14.95

Craig Sharp

Elite competitors experience diminishing returns from training: the better they become, the harder it is to improve. Yet an improvement of less than one per cent in, for example, swimming or athletics may well make the difference between a gold medal or nothing at all. As an extreme example, runner Said Aouita broke David Moorcroft's 5,000-metres world record of 13 minutes 0.41 seconds by 0.01 seconds — or 0.000013% (a margin less, presumably, than the accuracy of the measurement of the track).

Hence the unremitting search by coaches and competitors for techniques and substances to improve performance. Most use legal 'ergogenic aids', which range from nutritional supplements to swimmers' attempts to optimize their centre of buoyancy by sucking air into their rectums. But some use banned substances to modify their sports performance artificially, a practice officially known as 'doping'. Cynics suggest, with some justification, that the difference between ergogenic aids and doping is that doping works.

The astonishment of the sporting world when Seoul Olympic 100-metres winner Ben Johnson tested positive for anabolic steroids was not so much that he used them, but that he and his coach had apparently been so remarkably careless as to be detected. Contemporary explanations in the Olympic village ranged from his simply mistaking the tablets, to the shoulder-shrugging certainty expressed by a Cuban team doctor who told me he "knew" the CIA had spiked Johnson's drink, "because the Americans did not want a Canadian to win the 100 metres".

Barrie Houlihan, professor of sport policy at the University of Loughborough, sets his book on doping against this obsessively competitive background.

Most current books on doping concentrate on the pharmacology of the drugs and the technology of detection but, pleasingly, Houlihan takes much of this as read and pitches most of his book on the philosophical side. He defines the problem in terms of ethics and debates the evolution of anti-doping policies. A crucial aspect is the need to harmonize policies on testing methods and on sanctions or penalties — not only among the various sports governing bodies within each country, but also globally. But this issue is a lot easier to debate than to solve.

Rule 29 of the Olympic Charter begins



CORBIS/BETTMANN

Unfair game: Ben Johnson's winning Olympic medal was taken away when he tested positive for steroids.

with the simple sentence "Doping is forbidden". It is banned for two main reasons: sport is entirely an arbitrary activity, bounded by 'rules of the game', and taking drugs is as much a contravention of these rules as punching a soccer ball into the net with one's hand. It may also have adverse health effects.

Houlihan's discussion of the ethics involved is challenging and stimulating. His main point is that appealing to intuitive values is not enough. What is required is "the weight of democratic community condemnation and pervasive disapproval". Doping policies are never intellectually secure, in his view, but in need of continual "defence, support and refinement" based on education and resulting community support. These can act as a bulwark against cynical public acceptance of doping as a chemical version of the professional foul.

Houlihan explores the ethical and legal aspects of the urine-versus-blood controversy, and notes that Jehovah's Witnesses are one group that could object to venepuncture on religious grounds. It is not entirely facetious to predict a sudden increase in this faith among competitors, in line with the dramatic increase in apparent cardiac complaints in shooting teams at the 1984 Olympics, when β -blocking drugs were permitted if medically sanctioned. It may not have been entirely coincidental that β -blockers improve shooting performance; they are now on the International Olympic Committee's banned list.

Michele Verroken, director of ethics and anti-doping at the UK Sports Council, ends an excellent preface to the book with "After all, it is only a game!". That is the only statement of hers with which I disagree. World-

wide, sport is now a major industry, forming a large part of the entertainment industry. Participants are decreasingly club members and increasingly club employees. In Orwellian fashion, high-level sport is now work.

Overall, Houlihan has written a very readable, highly interesting and to some extent controversial book. It tackles the complex ethical, philosophical and policy problems head-on, and suggests more studies on the social reasons for doping. This book will greatly inform the often ill-informed debates on the topic, and will put into context the increasingly frequent legal cases in which disgruntled competitors who have tested positive sue their sport's governing bodies. □

Craig Sharp is in the Department of Sport Sciences, Brunel University, Osterley TW7 5DU, UK.

Peering into a child's priorities

The Nurture Assumption: Why Children Turn Out The Way They Do

by Judith Rich Harris

Bloomsbury/Free Press: 1998. 462 pp.

£18.99/\$26

Simon Baron-Cohen

Occasionally, a book is written about child development that is refreshing in breaking the mould. Judith Rich Harris's book is one of these. For a century at least, theories of child development have emphasized how the most important influence on a child is its parents (and in many theories, specifically its mother). Harris questions this sacred cow,

and in doing so, she opens up a space for a new theory not only of child development but also of cultural transmission. Her thesis is simple: the peer group is at least as important as parents in shaping child development, and there is plenty of evidence that it is more important.

At a first pass, this theory sounds crazy. Neonates have at least one parent, and for their first few years spend more time with a parent-figure than any peer group. So how can a peer group be having any significant effect on learning, personality, attitude formation or any other aspect of their mental development? Painstakingly, Harris takes the reader through some critical evidence — natural experiments, where the two rival theories can be teased apart and tested against each other. Here are some of her most persuasive cases.

Children who are born to immigrant parents, and who are raised in a language environment that is different from that spoken at home, do not end up speaking with their parents' immigrant accent, but instead effortlessly and rapidly acquire the accent of their peers. A moment's reflection brings home the truth of this elegant finding. We all know parents who continue to speak with a heavy 'foreign' accent, and who are therefore exposing their language-learning infant to this eccentric input — yet their children do not imitate these speech patterns. This is despite the fact that parents may reward their child with parental attention in conversation, and with love and affection. Rather, children systematically imitate the accent of their peers, with whom they may spend only a small proportion of their waking life.

Consider a second, equally compelling example. Babies born to profoundly deaf couples are exposed to sign language. This is not a needle-in-a-haystack situation — most deaf people marry other deaf people, but more than 90 per cent of the babies born to these couples have normal hearing. One might expect that the child would end up with little spoken language and with a preference for sign language. However, such children turn out to be as fluent speakers of the language of their peer group as children of hearing parents.

What is nice about these well-chosen examples is that the children can have only been learning the accent, or the spoken language, from one source — their peer group, not their parents. Moreover, the examples demonstrate that when a child has a choice over which type of input to prioritize, it tends to choose the influence of the peer group.

Harris's third example is again shockingly persuasive. Children born to parents who speak pidgin languages end up speaking something completely different: a creole. Thus, studies from Hawaii and elsewhere show that, under conditions where the labour force is drawn from many different



Child's play: peer pressure could mean anti-smoking campaigns are bound to fail with teenagers.

language groups, creating a sort of natural Tower of Babel situation, adults typically can communicate with one another only in a pidgin, which is a makeshift language lacking prepositions, articles, verb forms and a standardized word order. Each speaker of a pidgin speaks it a little differently, and the pidgin is often no more than a skimpy list of words that the speakers have in common. Yet the children of pidgin speakers speak a creole. A creole is a genuine language, with standardized word order and containing all of the linguistic properties that a pidgin lacks. A creole can express any idea, however abstract or complex, while a pidgin cannot. These children create their own language, which belongs to their peer group, and which ends up being more important to their social needs than either their parents' pidgin or ancestral tongue.

The final example is quite familiar. Deaf children put together in a school create not only their own community, but also their own sign language, even if this is discouraged by parents or adult teaching methods. Again, it is clear where the direction of influence is coming from, and what children are choosing to prioritize.

Harris is no narrow-minded theorist. She recognizes that parents have an enormous influence on their child's development, but she has audaciously challenged why, for decades, students of developmental psychology have been fed a diet of parent-child factors in the absence of an equally, if not more important, set of peer-group factors.

Her own history is interesting in this respect. She was rejected from the PhD pro-

gramme at Harvard, and so spent her adult life studying psychology on her own, as an outsider. She felt no pressure to conform to the dominant theories, but instead had the benefit of looking at the field from an impartial distance. The result is a new theory which I predict will stimulate many new research programmes. She finds her peer-group theory useful not just in explaining phenomena such as language acquisition, but also old chestnuts such as gender differences in behaviour, adolescent-parent conflicts and why anti-smoking campaigns are bound to fail with teenagers, to name but a few gems.

The book is long but very readable. Harris has an impressive breadth of knowledge, and entertainingly leads the reader from social development to genetics, from neuropsychology to criminology, and from social anthropology to linguistics and child-care. Even if her theory turns out to be wrong, which I doubt, I recommend this book to all students of developmental psychology as a stimulus for fresh thinking.

One final point: her book has helped me to make sense of a small empirical study of my own. Years ago, we investigated children with autism who are raised by parents where one or both have a different language from that of the peer group. We compared such children with their non-autistic siblings. Whereas the non-autistic siblings conformed exactly to Harris's predictions — they rapidly acquired the accent of their peers — the children with autism mostly acquired the accent of one of their parents, typically their mother (with whom they presumably had spent most time). The normal (non-

autistic) child has the strong need to be accepted by the peer group, to fit in and to belong. The case of autism starkly reminds us what happens when the normal brain circuits for social development malfunction, and how this affects not just the child's accent, but also their assimilation into culture. □

Simon Baron-Cohen is in the Departments of Experimental Psychology and Psychiatry, University of Cambridge, Cambridge CB2 3EB, UK.

A singularity in modern science

Geons, Black Holes, and Quantum Foam: A Life in Physics

by John Archibald Wheeler with Kenneth Ford

W. W. Norton: 1998. 380 pp. \$27.95, £19.95

Peter Galison

In the twentieth century theatre of physics, John Wheeler has stood at centre stage but always just out of the spotlight. He worked with Niels Bohr on the theory of nuclear fission in 1939 and on the Manhattan Project during the Second World War, and launched Richard Feynman on the quest that led him to quantum electrodynamics. Alongside Edward Teller, Wheeler helped to make thermonuclear weapons a reality in the late 1940s and early 1950s. From black holes to quantum measurement, from positronium to the collective model of the nucleus, Wheeler transformed an astonishing range of physics.

In a sadly conformist age, as herds of theorists thunder to one rumoured oasis after another, Wheeler has somehow maintained a quirky, intuitive, insightful style that is truly his own. There simply has been no one like him — he is and has always been a pragmatic visionary.

Pragmatic: Wheeler grew up an American boy who liked inventions and explosives. He was a theorist who, more than most of the other physicists at the wartime Metallurgical Lab in Chicago, ended up working well and learning easily from the DuPont engineers. Visionary: Bohr's institute in Copenhagen was about as far from American physics as it could be — a place where Bohr and his young associates agonized for days over getting the words, the physics and the philosophy right all at once. What comes through in *Geons, Black Holes, and Quantum Foam* is just how thoroughly, how improbably, how importantly Wheeler joined these American and European impulses.

Wheeler liked to work at the extremes. Could physics be done without fields — a world of particles alone? That became a long-standing project. But even that wasn't extreme enough. So Wheeler called up

Feynman one day in 1940 or 1941, and said (more or less), "Feynman, I know why all electrons and all positrons have the same charge." Why? "Because there's only one electron and it travels back and forth in time." In Feynman's hands this idea became a foundation of quantum electrodynamics in 1947–49. When the 'no fields' campaign faltered, Wheeler reversed course and fought on the opposite front, to reduce physics to fields without particles. When Wheeler wanted to explore the limits of gravitational physics, he took Robert Oppenheimer's 1939 speculations about black holes and pushed farther, doing physics at the extremity of the then quiescent field of general relativity.

Kenneth Ford has kept Wheeler's voice. It is a voice of striking honesty. There is none of the pretty glossing-over that so often

characterizes retrospective accounts of the Cold War, no pretending that everyone was on wonderful terms, that issues of national security were sidebars to the 'real world' of physics. No, Wheeler remembers Oppenheimer as a concatenation of brilliance and unstraightforward show-off. "My feelings toward him remain as they were more than 60 years ago," he says. "Oppenheimer was a complex human being. I never felt really close to him. I always felt I had to keep my guard up."

He speaks warmly of Teller, yet does not hesitate to criticize him. Nor does Wheeler ignore his own failings, such as his attachment to an idealized picture of Germany that may have delayed the launch of Los Alamos — without that delay, he believes, his brother might not have died on the battlefield. Wheeler says: "I was inclined to believe [as Werner Heisenberg did] that an immoral dictatorship was a transitory evil, something a

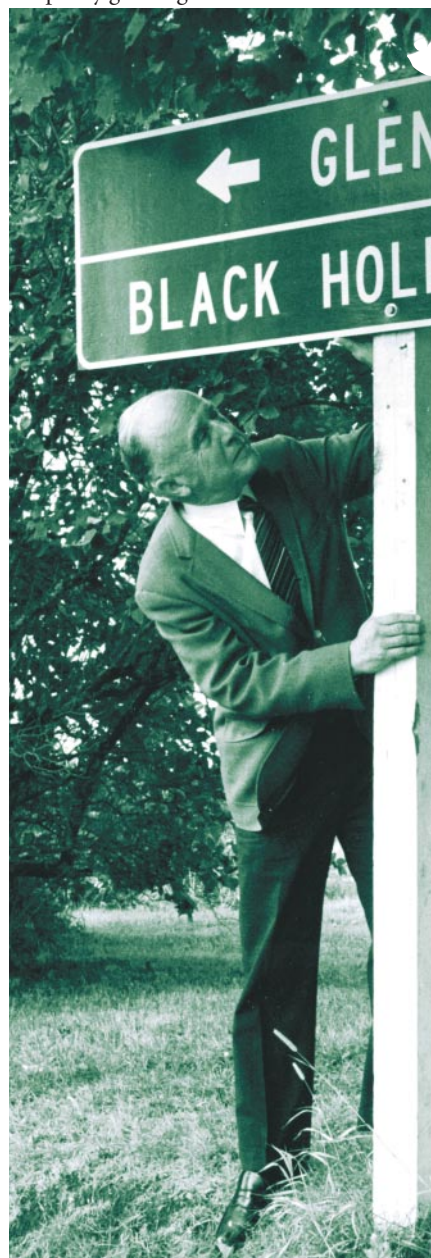
great nation could endure without lasting harm. Of course, I was wrong. So was Heisenberg, who never openly opposed the Nazi regime."

What is theoretical physics for Wheeler? It is a search to unify experience and at the same time to prosecute aesthetic concerns. "From the calculations and experiments that we call the nitty-gritty of our science to the most encompassing questions of philosophy, there is one unbroken chain of connection. There is no definable point along this chain where the truly curious physicist can say, 'I go only this far and no farther.'"

After the Second World War, most American physicists turned away from interpretative problems of quantum mechanics, shunting aside the great arguments launched by Bohr and Einstein in the 1930s over the meaning of measurement. Not Wheeler. Sometimes the goal was simply to augment understanding, as it was during his early involvement with Hugh Everett in the establishment of the 'many worlds' interpretation of quantum mechanics — a view that continues to attract attention from both physicists and philosophers.

Wheeler is a powerful singularity in twentieth-century physics. If this book helps remind us of that, it will have accomplished a great good thing. □

Peter Galison is in the Department of History of Science and Department of Physics, Science Center 235, Harvard University, Cambridge, Massachusetts 02138, USA.



Where next? Wheeler has stood at the crossroads of physics and explored diverse directions.

ROY BISHOP/AIP EMILIO SEGRE VISUAL ARCHIVES

Monsters in the night

Active Galactic Nuclei

by Julian H. Krolik
Princeton University Press: 1999. 598 pp.
\$39.50, £23.95 (pbk)

Chris Reynolds

A significant minority of galaxies possess a monster at their centre. These objects, known as active galactic nuclei (AGN), spew out vast quantities of radiation and rapidly moving material into space. Most astronomers now accept that AGN, one of the most extreme classes of objects in our Universe, are powered by the release of energy as material falls into the gigantic black hole that resides at the centre of each galaxy.

The plethora of terminology and concepts that have arisen in the field of AGN research means there is a desperate need for a comprehensive and up-to-date text. *Active Galactic Nuclei* by Julian Krolik fulfils that role. In varying levels of detail, Krolik treats almost every aspect of the AGN phenomenon, starting with the properties of the black hole itself and working outwards to the impact of AGN on their host galaxies. This global survey of AGN physics makes Krolik's book the most useful in the field for years.

Krolik's strength is his emphasis on the well-established physical principles underlying the phenomenology. The more speculative aspects of the field, such as the exotic physics occurring in the immediate vicinity of the black hole, are (justifiably) touched upon only briefly. Thorough descriptions of the fundamental physical processes punctuate the astronomical discussion. In particular, the description of the various mechanisms that generate and influence the observed radiation will be a valuable reference to both students and active AGN researchers. Throughout the book, the reader is given a flavour of the difficulties faced by observers, some of the modern observational breakthroughs and the interplay between the observational and theoretical sides of this vibrant branch of astrophysics.

A few parts of the discussion are confusing. For example, it might be difficult for anyone without a prior knowledge of stellar dynamics to follow the book's treatment of this subject. However, such sections are rare and detract only slightly from the overall strength of the book. One word of warning — the book assumes the reader to be fluent in college-level physics. The brief tutorials in the appendices (on general relativity, magneto-hydrodynamics and other important topics) will serve only to refresh a rusty mind on these topics and may be too brief for the non-specialist.

That said, this is an excellent text and a

valuable reference for anybody interested in the physics of these cosmic powerhouses. □
Chris Reynolds is at JILA, University of Colorado, Campus Box 440, Boulder, Colorado 80309, USA.

The charm of the single neuron

Biophysics of Computation: Information Processing in Single Neurons

by Christof Koch
Oxford University Press: 1998. 552 pp.
£45, \$59.95

Erik De Schutter

The double title of this book is slightly misleading, as Koch's monograph is mostly about the biophysics of single neurons. The biophysics are put in the context of computation, but little more should be expected here than division and multiplication, and readers are referred to other books for the information-processing side.

Instead, this book reviews models of the biophysical properties that allow neurons to process synaptic input and generate spikes. It focuses on the cellular level, including the passive and active properties of dendrites, spike initiation in realistic and simplified models of neurons and the dynamics of voltage-gated channels. Some subcellular aspects are reviewed also, but in an arbitrary fashion. For example, the role of calcium in learning at the postsynaptic side is described extensively, while its role at the presynaptic side is mentioned only briefly. Synaptic plasticity and quantal release are treated rather succinctly.

As the book positions itself between the black-box approaches used for neural-code analysis at the single-cell level and those used for the analysis of neural processing at the network level, it may serve as a bridge between these two fields. Christof Koch tries to give a basic neuroscience introduction to any given topic. Unfortunately, this is not done consistently. Many technical terms are not defined or not introduced at all — for instance, the reader is presumed to be familiar with "surround receptive fields" in visual cortex, or the "negative slope conductance" of the NMDA channel.

Moreover, the author has not respected the usual structure of a primer. Even in the introductory chapters he repeatedly deviates into a high-level discussion about related controversial issues; this can be of great interest to the specialized reader but may puzzle the undergraduate. In general, a critical mind is needed, as in some instances the numbers given for equation parameters are misleading or wrong. For example, the contribution of ions to the internal resistivity of cytoplasm is discussed in three different places, but in each case the values given are

valid for squid axons only, without mentioning this fact. A neuroscience background is required to assimilate most of the information in this monograph.

Unfortunately, as a primer about single-neuron biophysics it is rather unbalanced. In general, preference is given to previous work by the author and his collaborators. This leads to a detailed coverage of visual-cortex pyramidal cells and a few non-mammalian neurons, while other extensively investigated structures — such as the thalamic, auditory or olivocerebellar systems — are largely absent. Similarly, there is a separate chapter on bursting (which pyramidal neurons do), but mechanisms of oscillation (which they do not) are not treated systematically. Even in the proposed use of the pyramidal neuron as a canonical model, the book is not consistent, as realistic models of repetitive firing mechanisms are demonstrated by previous work by the author on the bullfrog sympathetic ganglion neuron.

This leads to a question that is surprisingly absent from this book: does a single canonical model suffice to describe neuronal computations, or may neurons be specialized to perform fundamentally different tasks? A chapter reviewing how differences in the morphology and expression of channels of particular neurons may correlate with their specific functions in the brain would have been a useful addition.

Although my review may seem rather critical, some of the chapters are excellent. For example, the chapter on the application of cable theory to neuronal dendrites contains the best treatment of transfer functions I have ever read. There are other gems like this in the book, which also has an excellent index. Unfortunately, it lacks the consistency and broadness to become a classic. □

Erik De Schutter is in the Department of Medicine, Born-Bunge Foundation, University of Antwerp-UIA, Universiteitplein 1, B2610 Antwerp, Belgium.

And on a higher order An Introduction to Natural Computation

by Dana H. Ballard
MIT Press, \$32.50, £19.95

Applied Neural Networks for Signal Processing

by Fa-Long Luo & Rolf Unbehauen
Cambridge University Press, £18.95, \$29.95

Brain Function and Oscillations. Volume I: Brain Oscillations. Principles and Approaches

by Erol Başar
Springer, £57, \$89.95

Rational Models of Cognition

edited by Mike Oaksford & Nick Chater
Oxford University Press, £60

Environmentally decoupled *sds*-wave Josephson junctions for quantum computing

Lev B. Ioffe^{*‡}, Vadim B. Geshkenbein^{†‡}, Mikhail V. Feigel'man[‡], Alban L. Fauchère[†] & Gianni Blatter[†]

^{*} Department of Physics and Astronomy, Rutgers University, Piscataway, New Jersey 08854, USA

[†] Theoretische Physik, ETH-Hönggerberg, CH-8093 Zürich, Switzerland

[‡] Landau Institute for Theoretical Physics, 117940 Moscow, Russia

Quantum computers have the potential to outperform their classical counterparts in a qualitative manner, as demonstrated by algorithms¹ which exploit the parallelism inherent in the time evolution of a quantum state. In quantum computers, the information is stored in arrays of quantum two-level systems (qubits), proposals for which include utilizing trapped atoms and photons^{2–4}, magnetic moments in molecules⁵ and various solid-state implementations^{6–10}. But the physical realization of qubits is challenging because useful quantum computers must overcome two conflicting difficulties: the computer must be scalable and controllable, yet remain almost completely detached from the environment during operation, in order to maximize the phase coherence time¹¹. Here we report a concept for a solid-state 'quiet' qubit that can be efficiently decoupled from the environment. It is based on macroscopic quantum coherent states in a superconducting quantum interference loop. Our two-level system is naturally bistable, requiring no external bias: the two basis states are characterized by different macroscopic phase drops across a Josephson junction, which may be switched with minimal external contact.

Our Josephson junction utilizes unconventional superconductors with order-parameter symmetry lower than the symmetry of the underlying crystal lattice. Recent phase-sensitive experiments on YBa₂Cu₃O₇ single crystals have established the *d*-wave nature of the copper oxide materials, thus identifying unambiguously the first unconventional superconductor^{12,13}. The sign change in the order parameter of these materials can be exploited to construct a new type of *s*-wave–*d*-wave–*s*-wave Josephson junctions exhibiting a degenerate ground state and a double-periodic current–phase characteristic. The basic idea is sketched in Fig. 1: connecting the positive (100) and negative (010) lobes of a *d*-wave superconductor with a *s*-wave material produces a π -loop with a current-carrying ground state characteristic of *d*-wave symmetry¹². Here we use an alternative geometry and match the *s*-wave superconductors (*S* in Fig. 1) to the (110) boundaries of the *d*-wave (*D*) material. The usual Josephson coupling (proportional to $(1 - \cos \phi)$) vanishes for symmetry reasons and we arrive at a bistable device, where the leading term in the coupling takes the form $E_d \cos 2\phi$ with minima

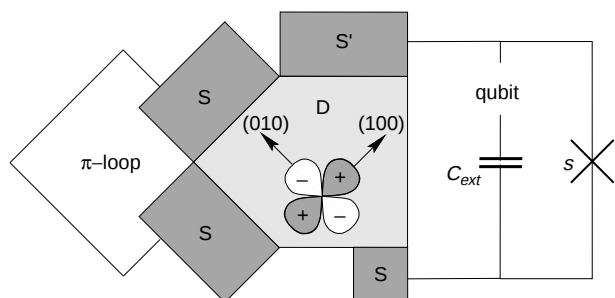


Figure 1 Schematic of the junction. Geometrical arrangements between *s*- and *d*-wave superconductors producing a π -loop (as in the phase-sensitive experiment by Wollman *et al.*¹²) and a qubit, basic building block of a quantum computer, are shown.

at $\phi = \pm \pi/2$ (here ϕ denotes the gauge-invariant phase drop across the junction and E_d is the coupling energy). In our design we need the minima at the positions $\phi = 0, \pi$ —the necessary shift is achieved by going over to an asymmetric SDS' junction with a large DS' coupling (Fig. 1). The static DS' junction shifts the minima of the active SD junction by the desired amount, $\pi/2$. A similar double-periodic junction was recently realized by combining two *d*-wave superconductors oriented at a 45° angle¹⁴.

The ground states of our SDS' junction are degenerate and carry no current, while still being distinguishable from one another: for example, after connecting the junction to a large inductance loop, the π state is easily identified through the induced current. It is this double-periodicity and the associated degeneracy in the ground state of the SDS' which we want to exploit here for quantum computation: combining the SDS' junction, a capacitor, and a conventional *s*-wave junction into a SDS' SQUID loop, we construct a bistable element that satisfies all the requirements for a qubit, the basic building block of a quantum computer. (SQUID indicates a superconducting quantum interference device.) Below we give a detailed account of the operational features of our device.

Consider a small-inductance (*L*) SQUID loop with $I_J L \ll \Phi_0$, where I_J denotes the (Josephson) critical current of the loop and $\Phi_0 = hc/2e$ is the flux unit. Such a loop cannot trap magnetic flux ($\Phi = 0$) and the gauge-invariant phase differences ϕ_1 and ϕ_2 across the two junctions follow each other, as the uniqueness of the

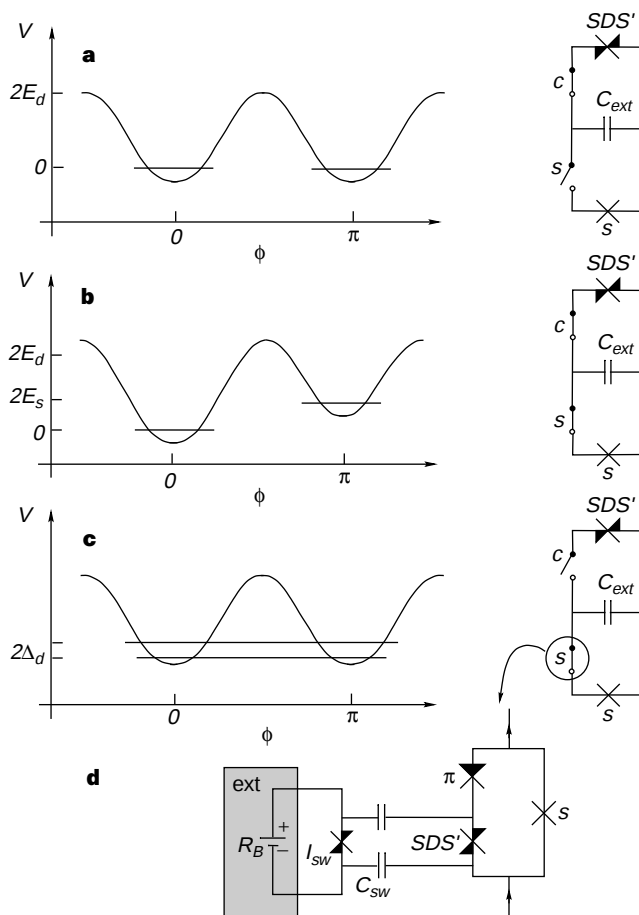


Figure 2 Energy-phase diagrams for the SDS' SQUID loop. **a**, Idle-state: the switches are set to 'c on' and 's off'—the relative dynamics are quenched, leaving the state unchanged. **b**, Phase-shifter: with the switch settings 'c on' and 's on' the relative phase between $|0\rangle$ and $|\pi\rangle$ increases linearly with time. **c**, Amplitude-shifter: the switch setting 'c off' isolates the *d*-wave junction. An initial state $|0\rangle$ oscillates back and forth between $|0\rangle$ and $|\pi\rangle$, allowing for a shift of amplitude. **d**, the SDS' junction, a π junction, and a *s*-wave junction combined into a SQUID loop and serving as a switch.

wavefunction requires that $\phi_1 - \phi_2 = 2\pi\Phi/\Phi_0$ (ref. 15). Combining an SDS' junction with a coupling energy E_d and a conventional s -wave junction (coupling E_s) into an SDS' SQUID loop, we obtain a potential energy

$$V(\phi) = E_d(1 - \cos 2\phi) + E_s(1 - \cos \phi) \quad (1)$$

exhibiting two minima at $\phi = 0, \pi$ (Fig. 2). The switch s allows us to manipulate their energy separation, choosing between minima which are either degenerate or separated by $2E_s$.

In the quantum case, the phase fluctuates as a consequence of the particle–phase duality¹⁵. The phase fluctuations are driven by the electrostatic energy required to move a Cooper pair across the junction, and are described by the kinetic energy $T(\dot{\phi}) = (\hbar/2e)^2 C \dot{\phi}^2/2$, where C denotes the loop capacitance. The dynamics of ϕ are manipulated by inserting a large switchable (by switch c) capacitance C_{ext} into the loop acting in parallel with the capacitances C_d and C_s of the d - and s -wave junctions. We note that the lagrangean $L = T - V$ of our loop is formally equivalent to that of a particle with ‘mass’ $m \propto C$ moving in the potential $V(\phi)$.

With the switch settings ‘ c on’ ‘ s off’ (Fig. 2a), the loop capacitance is large and the junction exhibits a double degenerate ground state which we characterize via the phase coordinated ϕ , $|0\rangle$ and $|\pi\rangle$. Closing the switch s (Fig. 2b), the degeneracy is lifted and while $|0\rangle$ becomes the new ground state, the $|\pi\rangle$ -state is shifted upwards by the energy $2E_s$ of the s -wave junction, the latter being frustrated (that is, pushed to maximal energy) when $\phi = \pi$. On the other hand, opening the switch c (Fig. 2c), completely isolates the d -wave junction and leads to the new ground and excited states $|\pm\rangle = [|0\rangle \pm |\pi\rangle]/\sqrt{2}$ separated by the tunnelling gap $2\Delta_d$. The latter relates to the barrier $2E_d$ and the capacitance C_d of the d -wave junction via¹⁵ $\Delta_d \propto E_d \exp(-2\sqrt{C_d E_d}/e^2)$. Closing the switch c , the capacitance is increased by C_{ext} and the tunnelling gap is exponentially suppressed. Using the above three settings, we can perform all the necessary single qubit operations, as follows.

Idle-state. The switch settings ‘ c -on’ and ‘ s -off’ define the qubit’s idle-state. While the large capacitance C_{ext} inhibits tunnelling, the degeneracy of $|0\rangle$ and $|\pi\rangle$ guarantees a parallel time evolution of the two states. This idle-state is superior to other designs, where the two states of the qubit have different energies and it is necessary to keep track of the relative phase accumulated between the basis states.

Phase shifter. Closing the switch s separates the energies of the basis states $|0\rangle$ and $|\pi\rangle$ by an amount $2E_s$. Using a spinor notation for the two-level system, the relative time evolution of the two states is described by the hamiltonian $H_s = -E_s \sigma_z$, with σ_z a Pauli matrix. Keeping the switch s on during the time t , the time evolution of the two states is given by the unitary rotation $u_z(\varphi) = \exp(-i\sigma_z \varphi/2)$ with $\varphi = -2E_s t/\hbar$.

Amplitude shifter. Assume we have prepared the loop in the ground state $|0\rangle$ and wish to produce a superposition by shifting some weight to the $|\pi\rangle$ state. Opening the switch c in the loop (Fig. 2c), the time evolution generated by the hamiltonian $H_d = \Delta_d \sigma_x$ of the open loop induces the rotation $u_x(\vartheta) = \exp(-i\sigma_x \vartheta/2)$ with $\vartheta = 2\Delta_d t/\hbar$. The system then oscillates back and forth between $|0\rangle$ and $|\pi\rangle$ with frequency $\omega = \Delta_d/\hbar$ and keeping the switch c open for an appropriate time interval t , we obtain the desired shift in amplitude (note that the qubit remains isolated from the environment during these Rabi oscillations).

Imposing the conditions $E_d \gg E_s$, Δ_d on the coupling energies, we make sure that the two states $|0\rangle$ and $|\pi\rangle$ are well defined while simultaneously involving only the low-energy states $|0\rangle$ and $|\pi\rangle$ of the system. Furthermore, all times involved should be smaller than the decoherence time τ_{dec} , requiring $E_s, \Delta_d \gg \hbar/\tau_{\text{dec}}$.

The present set-up differs significantly from the conventional (large-inductance) SQUID loop design, where the low-lying states are distinguished via the different amount of trapped flux and their manipulation involves external magnetic fields or biasing currents. SQUID loops of this type are being used in the design of classical

Josephson-junction computers¹⁶ and have also been proposed⁹ for the realization of quantum computers. However, this set-up suffers from the generic problem that the flux moving between the loops leads to a magnetic-field-mediated long-ranged interaction between the individual loops and also produces an unwanted coupling to the environment. By contrast, our device remains decoupled from the environment, the operating states do not involve currents, and switching between states can be triggered with a minimal contact to the external world—we therefore call our qubit implementation ‘quiet’.

We now consider how to perform two-qubit operations within an array of SDS' SQUID loops. A two-qubit state is a coherent superposition of single qubit states and can be expressed in the basis $\{|xy\rangle\}$, where $x, y \in \{0, \pi\}$ denote the phases on the d -wave junctions of the first (x) and second (y) qubit, respectively. Unitary operations acting on these states are represented as 4×4 unitary matrices. Single-qubit operations u acting on the second qubit take the block-matrix form

$$U_2 = \begin{pmatrix} u & 0 \\ 0 & u \end{pmatrix} \quad (2)$$

and a similar block form selecting odd and even rows and columns defines the single-qubit operations on the first qubit. As all logic operations on two qubits can be constructed from combinations of single-qubit operations and the ‘controlled-NOT’ gate¹ it is sufficient to define the operational realization of the latter. The controlled-NOT gate performs the following action on the two qubits: with the first (controller) qubit in state $|x\rangle$ and the second (target qubit) in state $|y\rangle$ the operation shall leave the target qubit unchanged if $x = 0$, while flipping it between 0 and π when $x = \pi$, in matrix notation:

$$U_{\text{CNOT}} = \begin{pmatrix} 1 & 0 \\ 0 & \sigma_x \end{pmatrix} \quad (3)$$

The above controlled-NOT operation can be constructed from the 2-qubit phase shifter. Connecting two individual qubits in their idle-state over a s -wave junction into a SQUID loop, the states $|00\rangle$ and $|\pi\pi\rangle$ become separated from the states $|0\pi\rangle$ and $|\pi 0\rangle$ by the energy $2E_{s_b}$ of the s -wave junction. Keeping the two qubits connected during the time t introduces a phase shift $\chi = -2E_{s_b} t/\hbar$ between the two pairs of states:

$$U_{\text{ps}}(\chi) = \begin{pmatrix} u_z(\chi) & 0 \\ 0 & u_z(-\chi) \end{pmatrix} \quad (4)$$

The controlled-NOT gate (equation (3)) then can be constructed from the phase-shifter (equation (4)) via the following sequence of single- and two-qubit operations (see ref. 6 for a similar realization of the controlled-NOT gate)

$$U_{\text{CNOT}} = \exp(-i\pi/4) U_{2y}(\pi/2) U_{1z}(-\pi/2) U_{2z}(-\pi/2) \cdot U_{\text{ps}}(\pi/2) U_{2y}(-\pi/2) \quad (5)$$

where the single qubit operations $U_{i\mu}(\theta)$ rotate the qubit i by an angle θ around the axis μ ($u_{i\mu}(\theta) = \exp(-i\sigma_{i\mu} \theta/2)$ acting on i) while leaving the other qubit unchanged.

The switches are important elements in our design, and a valid suggestion for their implementation is the single electron transistor¹⁷. Here we propose a quiet switch design optimally adapted to our SDS' qubits. The basic idea derives from frustrating the junctions in a SQUID loop, resulting in a ‘phase blockade’: combining an SDS' junction with energy E_d , a π -junction with $E_\pi \ll E_d$, and an s -wave junction with $E_s = E_\pi$ into a (small-inductance) SQUID loop (Fig. 2d), we obtain the following switching behaviour. The phase $\phi = 0$ on the SDS' junction frustrates the remaining junctions, the loop’s energy–phase relation is a constant, $E_{\text{sw}}(\phi_\pi = \phi_s - \pi) \equiv 0$, and the switch is open. A voltage pulse coming down the signal lines and switching the SDS' junction into the $|\pi\rangle$ state changes the phase relation between

the π - and the s -wave junctions and closes the switch: the energy $E_{sw}(\phi_\pi = \phi_s) = 2E_\pi(1 - \cos \phi_\pi)$ produces the current–phase relation $I = (2e/\hbar)\partial_{\phi_\pi} E_{sw}$, thus suppressing the fluctuations of the phase across the switch. The appropriate voltage pulses could be generated by driving an external SDS' junction unstable.

The quiet-device concept proposed above relies heavily on the double periodicity of the SD junction. As the second harmonic is strongly suppressed in a SID (s -wave–insulator– d -wave) tunnel junction, a more feasible suggestion for the realization of a $\cos 2\phi$ junction is the SND 'sandwich', where the superconductors are separated by a thin metallic layer N. For a clean metallic layer, the coupling energies for the n th harmonic are large and of the order of $E_l \approx k_F^2 A \hbar v_F / d$, producing the well known saw-tooth shape in the current–phase relation¹⁸ (here, v_F (k_F) denotes the Fermi velocity (wavenumber) in the N layer while d and A are its width and area). In reality, it seems difficult to deposit a clean metallic film on top of a d -wave superconductor, and we have to account for the reduction in the coupling E_l due to the finite scattering length l in the metal layer. Using quasi-classical techniques to describe a dirty SN_dD junction (s -wave–dirty normal metal– d -wave), we obtain a second-harmonic coupling energy $E_d \approx k_F^2 A (\hbar v_F / d) (l/d)^3 \approx (R_Q/R) (l/d) E_T$, where l denotes the scattering length in the normal metal, $R_Q = \hbar/e^2$ is the quantum resistance, R is the junction resistance, and $E_T = (\hbar v_F / d) (l/d)$ is the Thouless energy.

The second important device parameter is the tunnelling gap Δ_d , which depends quite sensitively on the coupling to the environment. The usual reduction in the tunnelling probability by the environment¹⁹ is modified if the system is effectively gapped at low energies²⁰. This is the case for our $SN_dDN'_dS'$ junction where the low-energy quasiparticle excitations in the metal are gapped over the Thouless energy E_T (ref. 21). The dynamics of the junction are only affected by the presence of virtual processes involving energies larger than E_T , leading to a renormalized capacitance $C_{ren} \approx \hbar/RE_T$ (compare ref. 20) and resulting in the reduced tunnelling gap $\Delta_d \propto E_d \exp[-\nu(R_Q/R)\sqrt{l/d}]$, with ν of the order of unity. Consistency requires that the tunnelling process is 'massive' and hence slow, $\hbar/\tau < E_T$. With a tunnelling time $\tau \approx S/E_d$ ($S \approx \hbar(R_Q/R)\sqrt{l/d}$ gives tunnelling action) we find that the constraint $\hbar/\tau E_T \approx \sqrt{l/d} < 1$ is satisfied. The condition $\Delta_d \ll E_d$ requires the tunnelling gap Δ_d to be small, but large enough in order to allow for reasonable switching times, requiring $(R_Q/R)\sqrt{l/d}$ to be of the order of 10. With typical device dimensions $d \approx 1,000$ Å, $l \approx 10$ Å, and $R/R_Q \approx (d/l)(1/Ak_F^2) \approx 1/100$, this condition can be realized. The operating temperature T is limited by the constraint $S/\hbar > E_d/T$, guaranteeing that our device operates in the quantum regime, and the requirement $T < E_T$, that thermal quasi-particle excitations be absent. The first condition takes the form $T \ll \hbar/\tau \approx \sqrt{l/d} E_T$ and is the more stringent. Using the above parameters and a typical value $v_F \approx 10^8$ cm s⁻¹, we obtain a Thouless energy $E_T \approx 1$ K and hence $T \ll 0.1$ K.

An important topic in quantum computation is decoherence. Within our set-up it is the Thouless gap E_T which inhibits the excitation of quasiparticles. The longest trajectories in the $SN_dDN'_dS'$ junction limiting the size of E_T are those connecting the two s -wave superconductors over a path of length $L_{max} \approx 2d^2/l + d_d$, where the first term stems from the diffusive propagation through the two dirty metal layers, and the second term is due to the ballistic trajectory along the node of the d -wave superconductor (of width d_d). The conditions $\xi_d \ll d_d \ll d^2/l$, with ξ_d the coherence length in the d -wave superconductor, then guarantee that the d -wave layer is thick enough to avoid the quenching through the proximity to the s -wave superconductors, while still being thin enough to leave the Thouless gap unchanged. A further source of decoherence are the switches. Within the set-up sketched in Fig. 2d, the voltage pulse triggering the SDS' junction in the switch is generated through an external junction driven by a current source. We then have to make sure that while the switching signal reaches the SDS' junction of the

switch loop, the external noise is kept away from the qubit's SDS' junction. Most dangerous is the low-frequency part of the noise spectrum, as it would induce brownian motion of the relative phase in the wavefunction of the qubit. An appropriate filtering can be achieved through a capacitive decoupling (C_{sw}) of the external driven junction generating the pulses and the SDS' switch, transparent at the high frequencies of the triggering pulse, but suppressing the low-frequency noise of the current source through a factor $(\omega C_{sw} R_B)^2$, where R_B denotes the resistivity of the external current source. Furthermore, the driven junction loop acts as a shunt, producing an additional suppression factor $(e\omega/I_{sw})^2 R_Q/R_B$. □

Received 14 December 1998; accepted 8 March 1999.

- Ekert, A. & Jozsa, R. Quantum computation and Shor's factoring algorithm. *Rev. Mod. Phys.* **68**, 733–753 (1996).
- Cirac, J. I. & Zoller, P. Quantum computations with cold trapped ions. *Phys. Rev. Lett.* **74**, 4091–4094 (1995).
- Monroe, C., Meekhof, D., King, B., Itano, W. & Wineland, D. Demonstration of a fundamental quantum logic gate. *Phys. Rev. Lett.* **75**, 4714–4717 (1995).
- Turchette, Q., Hood, C., Lange, W., Mabushi, H. & Kimble, H. J. Measurement of conditional phase shifts for quantum logics. *Phys. Rev. Lett.* **75**, 4710–4713 (1995).
- Gershenfeld, N. A. & Chuang, I. L. Bulk spin-resonance quantum computation. *Science* **275**, 350–356 (1997).
- Loss, D. & DiVincenzo, D. P. Quantum computation with quantum dots. *Phys. Rev. A* **57**, 120–126 (1998).
- Shnirman, A., Schön, G. & Hermon, Z. Quantum manipulations of small Josephson junctions. *Phys. Rev. Lett.* **79**, 2371–2374 (1997).
- Averin, D. V. Adiabatic quantum computation with Cooper pairs. *Solid State Commun.* **105**, 659–664 (1998).
- Bocko, M. F., Herr, A. M. & Feldman, M. J. Prospects for quantum coherent computation using superconducting electronics. *IEEE Trans. Appl. Supercond.* **7**, 3638–3641 (1997).
- Kane, B. E. A silicon-based nuclear spin quantum computer. *Nature* **393**, 133–137 (1998).
- Haroche, S. & Raimond, J.-M. Quantum computing: dream or nightmare? *Phys. Today* **49**, 51–52 (1996).
- Wollman, D. A., Van Harlingen, D. J., Lee, W. C., Ginsberg, D. M. & Leggett, A. J. Experimental determination of the superconducting pairing state in YBCO from the phase coherence of YBCO-Pb DC SQUID's. *Phys. Rev. Lett.* **71**, 2134–2137 (1993).
- Kirtley, J. R. *et al.* Symmetry of the order parameter in the high- T_c superconductor $YBa_2Cu_3O_{7-x}$. *Nature* **373**, 225–228 (1995).
- Il'ichev, E. *et al.* Anomalous periodicity of the current-phase relationship of grain-boundary Josephson junctions in high- T_c superconductors. Preprint cond-mat/9811017 at (<http://xxx.lanl.gov>) (1998).
- Tinkham, M. in *Introduction to Superconductivity* Ch. 6.4 and 7.3, (McGraw-Hill, Singapore, 1996).
- Likharev, K. K. & Semenov, V. K. RSFQ logic/memory family: A new Josephson-junction technology for sub-terahertz-clock-frequency digital systems. *IEEE Trans. Appl. Supercond.* **1**, 3–28 (1991).
- Joyez, P., Lafarge, P., Filipe, A., Esteve, D. & Devoret, M. H. Observation of parity-induced suppression of Josephson tunneling in the superconducting single electron transistor. *Phys. Rev. Lett.* **72**, 2458–2461 (1994).
- Ishii, C. Josephson currents through junctions with normal metal barriers. *Prog. Theor. Phys.* **44**, 1525–1547 (1970).
- Caldeira, A. O. & Leggett, A. J. Quantum tunneling in a dissipative system. *Ann. Phys.* **149**, 374–456 (1983).
- Ambeagaokar, V., Eckern, U. & Schön, G. Quantum dynamics of tunneling between superconductors. *Phys. Rev. Lett.* **48**, 1745–1748 (1982).
- Golubov, A. A. & Kuprianov, M. Yu. Theoretical investigation of Josephson tunnel junctions with spatially inhomogeneous superconducting electrodes. *J. Low Temp. Phys.* **70**, 83–130 (1988).

Acknowledgements. We thank A. Kitaev, D. Loss, A. Millis and B. Spivak for discussions. This work was supported by the Fonds National Suisse.

Correspondence and requests for materials should be addressed to G.B. (e-mail: blatterj@itp.phys.ethz.ch).

Metastable ice VII at low temperature and ambient pressure

S. Klotz*, J. M. Besson*, G. Hamel†, R. J. Nelmes‡, J. S. Loveday‡ & W. G. Marshall§

* *Physique des Milieux Condensés UMR 7602 and † Département des Hautes Pressions, Université P&M Curie, B77, 75252 Paris, France*

‡ *Department of Physics and Astronomy, The University of Edinburgh, Edinburgh EH9 3JZ, UK*

§ *ISIS Facility, Rutherford Appleton Laboratory, Chilton, Didcot OX11 0QX, UK*

Ice exhibits many solid-state transformations under pressure, and also displays a variety of metastable phases¹. Most of the high-pressure phases of ice can be recovered at ambient pressure provided that they are first cooled below about 100 K. These ice polymorphs might exist on the surfaces of several satellites of the

outer planets². One of the few exceptions to this (meta)stability on quenching has been ice VII, the dominant high-pressure phase. Here we show that isothermal compression of D₂O ice VI below 95 K produces pure ice VII, and that this phase can remain stable at atmospheric pressure. It remains metastable indefinitely at 77 K. Like the other recoverable ice phases, it transforms to low-density amorphous ice between about 120 and 150 K at 1 bar. The temperature range over which ice VII remains metastable increases markedly on compression to 6 GPa, indicating that ice VII is in fact the most robust of all the metastable ice phases.

Although all attempts to produce ice VII at low temperatures by cooling it very rapidly have failed so far (ice VII transforms to phase VIII below ~269 K, even when the temperature is decreased very rapidly at cooling rates of ~20 K s⁻¹ (A. I. Kolesnikov, personal communication)) there is recent experimental evidence that ice-VII-like structures can be found outside the equilibrium field of phase VII. Hemley *et al.*³ realized that Raman spectra of ice I_h compressed between 77 and 150 K to beyond 4 GPa showed some resemblance to those of ordinary ice VII at high temperatures; they denoted this as an ice VII' phase. In the first study of the structural details of such cold-compressed ice samples, Finney *et al.*^{4,5} recorded neutron diffraction spectra from samples obtained by compressing phases II and VI to 4 GPa at 120 K, and confirmed the resemblance to ice VII. However, these spectra contained some non-ice-VII features and it remained uncertain whether the phase produced was true ice VII, and could be retrieved at ambient pressure.

We have now found that, by cooling ice VI further and compressing it below 95 K, pure ice VII may be obtained. Neutron diffraction data were collected at the PEARL/HiPr and POLARIS stations of the ISIS facility. Water (D₂O) was loaded into a modified Paris-Edinburgh high-pressure cell with helium as the hydraulic fluid in the ram, compressed to 1.5 GPa at 300 K into phase VI, and then cooled to 95 K. Ice VI compresses completely regularly up to 3.8 GPa, where an abrupt transition occurs which is easily observable by a sudden loss in load of 5%. Figure 1a shows a diffraction

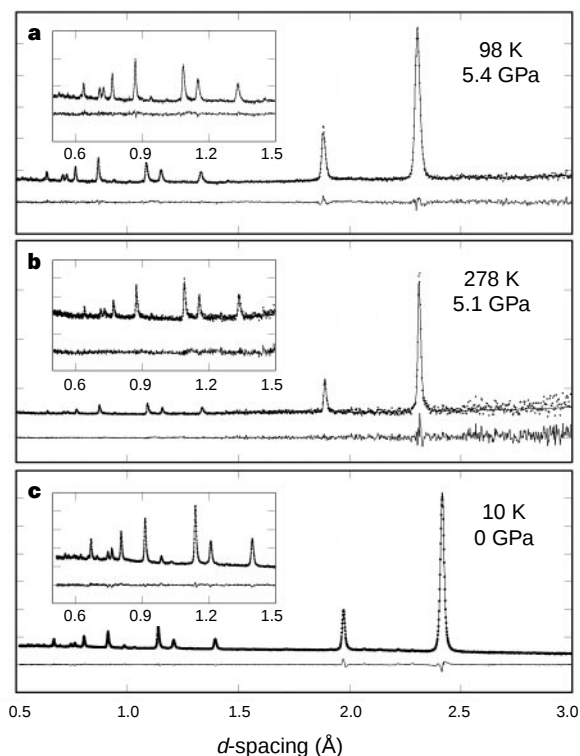


Figure 1 Neutron diffraction patterns of ice VII under different conditions. The lines through the data points are fits by Rietveld refinements, and the difference curves are given below each spectrum. The insets show the short d-spacing range enlarged.

Table 1 Refined structural parameters of D₂O ice VII

	V_m (cm ³ mol ⁻¹)	$U_{iso}(O)$ (10 ⁻² Å ²)	O–D (Å)
278 K, 5.1 GPa	10.541(4)	1.99(8)	0.933(3)
98 K, 5.4 GPa	10.310(3)	1.17(5)	0.931(2)
98 K, 0 GPa*	12.108(4)	1.71(8)	0.905(2)
10 K, 0 GPa†	12.05(2)	1.54(4)	0.911(1)

Space group, $Pn3m$; $Z = 2$. V_m is the molar volume, $U_{iso}(O)$ is the isotropic thermal parameter of the oxygen atom, and O–D the apparent covalent O–D bond length. Errors are given in parentheses.

* In high-pressure cell.

† In helium cryostat.

pattern obtained from this phase at 5.4 GPa and 98 K; Fig. 1b shows a spectrum of ice VII obtained at 278 K at approximately the same pressure. It can be seen that the fit of the former to the ice VII structure is excellent, and that there are no non-ice-VII features. All refinements were done assuming the known cubic structure (space group $Pn3m$) with the oxygen at the 2a site (1/4, 1/4, 1/4) and the deuterium at (x, x, x), where $x \approx 0.42$. They give molar volumes, apparent O–D distances and thermal parameters as listed in Table 1. A comparison with the values tabulated⁶ for ice VII at 278 K shows excellent agreement apart from the (expected) temperature difference in the isotropic thermal parameter of the oxygen atom, $U_{iso}(O)$. We conclude that the phase obtained by compressing ice VI below 95 K is (metastable) ice VII.

The sample was then decompressed at ~100 K to ambient pressure, and diffraction data were taken with the sample still in the cell (Table 1). Some significant changes in $U_{iso}(O)$ and the apparent O–D distance were found—as discussed further below—but the sample remained pure ice VII. Several pellets of ice VII were produced in the same way and one was transferred at 77 K to a vanadium can which was inserted into a He cryostat. This procedure ensured that the sample was definitely at ambient pressure and allowed cooling to 4 K. Figure 1c shows a spectrum collected at 10 K for ~10 hours. Again it is pure (metastable) ice VII. (The ambient-pressure spectrum reveals a residual peak width ~50% larger than that of ice VII at 300 K. According to the refined profile coefficients, this mainly reflects the presence of ± 0.05 GPa strain with a particle size not smaller than 0.1 μ m, which is comparatively large in view of the abrupt fabrication conditions at low temperatures.) One sample was stored at 77 K for six months and subsequent neutron diffraction data showed no evidence of any changes during this period.

We have also determined the structural pressure dependence at

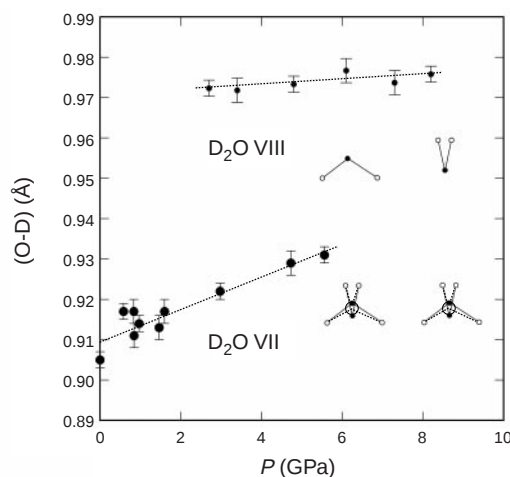


Figure 2 Pressure dependence of the refined covalent O–D distance in ices VII and VIII (refs. 7, 8). The inset drawings illustrate the orientation of two neighbouring D₂O molecules, and show how the O-site displacements in the disordered phase (bottom pair of drawings) cause the apparent distance (dotted lines) to be shorter than the actual value (solid lines) found in ordered ice VIII (top pair of drawings).

98 K on pressure release. The most interesting result concerns the refined values shown in Fig. 2 for the covalent O–D bond length which give an approximately linear variation ($\partial(\text{O–D})/\partial P \approx +0.004 \text{ \AA GPa}^{-1}$) up to 5.4 GPa, including the 0–2 GPa range inaccessible at 300 K. This is an order of magnitude larger than $\partial(\text{O–D})/\partial P$ in ice VIII (refs 7, 8), also shown in Fig. 2. Moreover, the (apparent) O–D distances are significantly smaller than in ice VIII. The latter effect arises from a disordering of oxygen atoms over sites displaced by $\delta \approx 0.1 \text{ \AA}$ from the average position in ice VII (ref. 9). This results in an apparent bond-length shortening of $0.05 \text{ \AA}$, as shown by the insets in Fig. 2, and correspondingly increased oxygen thermal parameters $U_{\text{iso}}(\text{O})$. The changes in O–D and $U_{\text{iso}}(\text{O})$ between 5.4 and 0 GPa at 98 K (Table 1) can thus be interpreted as a significant increase in δ over this range.

The molar volume of ice VII at ambient pressure and 98 K is found to be $12.108 (\pm 0.004) \text{ cm}^3 \text{ mol}^{-1}$ (Table 1). This is within error the same as that found for ice VIII at the same pressure and temperature¹⁰, and only 0.4% off the value obtained by extrapolation of the 300 K equation of state⁸. Thus, the equation of state of ice VII at low temperatures and pressures is completely regular and—as found for the classical VII–VIII transition at 270 K and $\sim 2 \text{ GPa}$ —the molar volume of ices VII and VIII are identical within experimental error. Hence, the disordering of the molecular orientation in recovered ice VII does not measurably change the volume of the overall structure. However, the disorder in low-temperature ice VII cannot be dynamical, as it is above 270 K, because molecular reorientation is known to be negligible at $\sim 100 \text{ K}$ and below¹.

To explore the stability limits of low-temperature ice VII, the sample temperature was gradually raised above 100 K (Fig. 3). The first change observed was the appearance at $\sim 115 \text{ K}$ of the broad features marked by arrows in trace a and the left inset. They correspond closely to the positions of the strongest additional reflections of ice VIII, begin to decrease in intensity at 125 K, and disappear by 135 K. The ice VII peaks start to decrease in intensity at 120 K, and a broad feature at $\sim 3.7 \text{ \AA}$ starts to grow at 125 K. A sequence of spectra collected at 138 K with 20-minute accumulation time showed a gradual disappearance of all ice VII diffraction peaks,

which was completed after ~ 1.5 hours, as illustrated in trace b. To avoid any possible further transformations, the temperature was then decreased to 120 K and the spectrum shown in the right inset was collected. The pattern is dominated by scattering from an amorphous form, but several weak but sharp (crystalline) features are visible, as marked by numbers 1–6. Peaks 4–6 may be scattering from the vanadium of the cryostat and sample can, but peaks 1–3 are definitely not, and peaks 1 and 2 have been observed previously in samples heated from recovered ice VIII (ref. 11).

The sample was then heated back above 138 K and transformed at 152 K to a phase with a spectrum shown in Fig. 3 trace c. It can be unambiguously identified as that of cubic ice I_c by comparison with published neutron data¹¹. On further heating, the sample transformed to hexagonal I_h (trace d), in accord with the known behaviour of ice I_c .

The stability limit of recovered ice VII was also explored as a function of pressure. A sample heated isobarically at 1.1 GPa showed a sluggish transition to ice VIII, which starts at $\sim 135 \text{ K}$ and is completed at 180 K. Another sample heated isobarically at 6 GPa showed a similarly sluggish transition to ice VIII starting at $\sim 180 \text{ K}$ and ending at $\sim 210 \text{ K}$. The range of metastability of low-temperature ice VII seems therefore to increase considerably with pressure. We note that the Raman spectra of Hemley *et al.*'s samples do not indicate any further transformation up to at least 30 GPa (ref. 3). Therefore, it seems very likely that ice VII remains metastable over most of the domain of the thermodynamically stable phase ice VIII. Thus low-temperature ice VII emerges as the most 'robust' metastable ice phase yet discovered.

The transformations that we have found in recovered ice VII have much in common with those found in recovered ice VIII by Balagurov *et al.*¹¹. Both transform to an amorphous phase, and the diffraction data given in ref. 11 show a strong resemblance to trace b in Fig. 3, including the sharp feature at $3.0 \text{ \AA}$ marked as peak 2 in the right inset. A comparison with structure factor data of high- and low-density amorphous (HDA and LDA) ices, shown in Fig. 3, right inset, reveals that recovered ice VII transforms into LDA. Our data reproduce remarkably well all the characteristic features of LDA, and rule out any presence of HDA. Contrary to Balagurov *et al.*'s interpretation of their data, we suggest that the same is true for ice VIII. (The existence of HDA in the data of ref. 11 relies entirely on the weak feature at $3.0 \text{ \AA}$. This peak, which is identical with peak 2 of our data (Fig. 3, right inset), is sharp and hence not attributable to HDA but to a minority crystalline phase. Also, HDA is known to be unstable beyond $\sim 125 \text{ K}$, where it transforms irreversibly to LDA¹².) An important difference seems to be that recovered ice VIII was found to be stable up to the slightly higher temperature of 130 K (ref. 11). We note that the ice-VIII-like features marked by arrows in Fig. 3 appear at ~ 115 – 120 K —consistent with a linear extrapolation of the pressure dependence of the ice VII–VIII transition temperature (see above)—and disappear at the same temperature as for ice VIII (130–135 K)¹¹. This is in accordance with the identification of these features as the formation of incipient or nanocrystalline ice VIII, and supports the apparent 10 K difference in stability range of recovered ice VII and ice VIII.

Thus the high-pressure, high-temperature form of ice can be obtained at low temperatures and ambient pressure—not by recovery from its equilibrium field but through a non-equilibrium transition from ice VI—and is metastable over a large range of pressure. The availability of ice VII in gram quantities at ambient pressure opens up new possibilities for its detailed characterization. It also becomes possible to make comparative studies of ices VIII and VII. These constitute an important pair, in that they differ principally in their proton ordering; also, with their two interpenetrating H-bonded networks, they are a valuable complementary system to the single-network ices XI and I_h —the only other ordered-disordered pair which can be obtained at ambient pressure. □

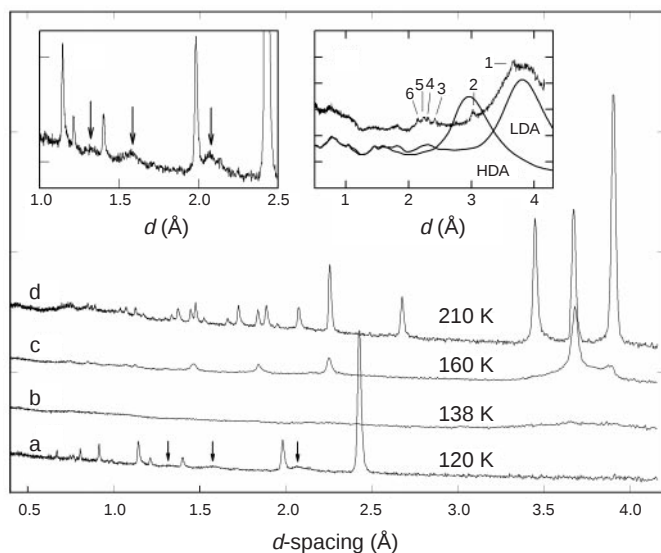


Figure 3 Neutron diffraction spectra showing the transition of ice VII at ambient pressure. Arrows in spectrum a mark weak broad features appearing at 120 K, and the left-hand inset shows an enlarged part of this pattern. The right-hand inset shows a spectrum of the amorphous phase at 120 K after cooling from 138 K (spectrum b), compared to published data¹³ on the structure factors of HDA and LDA ices. (Values for q cited in ref. 13 were shifted by 2.5% to account for slight discrepancies with our data as well as with the X-ray results of refs 14, 15.) Numbers label weak features of a crystalline phase. Spectra c and d correspond to ice phases I_c and I_h , respectively.

Received 16 October 1998; accepted 8 February 1999.

1. Hobbs, P. V. *Ice Physics* (Clarendon, Oxford, 1974).
2. Gaffney, E. S. Water ice polymorphs and their significance on planetary surfaces. *Icarus* **44**, 511–519 (1980).
3. Hemley, R. J., Chen, L. C. & Mao, H. K. New transformations between crystalline and amorphous ice. *Nature* **338**, 638–640 (1989).
4. Finney, J. L., Lobban, C., Klotz, S. & Besson, J. M. in *Collected Abstracts of the XVII Congress and General Assembly* (ed. Griffin, J. F.) C-534 (IUCr, Seattle, 1996).
5. Finney, J. L., Lobban, C., Klotz, S., Besson, J. M. & Hamel, G. in *ISIS 96—Rutherford Appleton Laboratory Report RAL 96-050* (ed. Taylor, A. D.) Vol. A17 (Chilton, UK, 1996).
6. Kuhs, W. F., Finney, J. L., Veltier, C. & Bliss, D. V. Structure and hydrogen ordering of ices VI, VII, and VIII by neutron powder diffraction. *J. Chem. Phys.* **81**, 3612–3623 (1984).
7. Nemes, R. J. *et al.* Neutron diffraction study of the structure of deuterated ice VIII to 10 GPa. *Phys. Rev. Lett.* **71**, 1192–1195 (1993).
8. Besson, J. M. *et al.* Variation of interatomic distances in ice VIII to 10 GPa. *Phys. Rev. B* **49**, 12540–12550 (1994).
9. Nemes, R. J. *et al.* Multi-site disordered structure of ice VII to 20 GPa. *Phys. Rev. Lett.* **81**, 2719–2722 (1998).
10. Besson, J. M. *et al.* Structural instability in ice VIII under pressure. *Phys. Rev. Lett.* **78**, 3141–3144 (1997).
11. Balagurov, A. M. *et al.* Neutron-diffraction study of phase transitions of high-pressure metastable ice VIII. *JETP Lett.* **53**, 30–34 (1991).
12. Mishima, O., Calvert, L. D. & Whalley, E. *Nature* **314**, 76–78 (1985).
13. Bellissent-Funel, M.-C., Teixeira, J. & Bosio, L. Structure of high-density amorphous water. II. Neutron scattering study. *J. Chem. Phys.* **87**, 2231–2235 (1987).
14. Bizid, A., Bosio, O., Defrain, A. & Oumezzine, M. Structure of high-density amorphous water. I. X-ray diffraction study. *J. Chem. Phys.* **87**, 2225–2230 (1987).
15. Bosio, L., Johari, G. P. & Teixeira, J. X-ray study of high-density amorphous water. *Phys. Rev. Lett.* **56**, 460–463 (1986).

Acknowledgements. This work was supported by the UK EPSRC and the UK CCLRC, as well as the French Institut National des Sciences de l'Univers and the French Commissariat à l'Energie Atomique.

Correspondence and requests for materials should be addressed to S.K. (e-mail: stk@pmc.jussieu.fr).

Low-temperature superplasticity in nanostructured nickel and metal alloys

S. X. McFadden*, R. S. Mishra*, R. Z. Valiev†, A. P. Zhilyaev† & A. K. Mukherjee*

* Department of Chemical Engineering and Materials Science, University of California, Davis, California 95616, USA

† Institute of Physics and Advanced Materials, Ufa State Aviation Technical University, 12, K. Marx Street, Ufa, 450000 Russia

Superplasticity—the ability of a material to sustain large plastic deformation—has been demonstrated in a number of metallic, intermetallic and ceramic systems. Conditions considered necessary for superplasticity¹ are a stable fine-grained microstructure and a temperature higher than $0.5 T_m$ (where T_m is the melting point of the matrix). Superplastic behaviour is of industrial interest, as it forms the basis of a fabrication method that can be used to produce components having complex shapes from materials that are hard to machine, such as metal matrix composites and intermetallics. Use of superplastic forming may become even more widespread if lower deformation temperatures can be attained. Here we present observations of low-temperature superplasticity in nanocrystalline nickel, a nanocrystalline aluminium alloy (1420-Al), and nanocrystalline nickel aluminide (Ni_3Al). The nanocrystalline nickel was found to be superplastic at a temperature 470°C below that previously attained²: this corresponds to $0.36 T_m$, the lowest normalized superplastic temperature reported for any crystalline material. The nanocrystalline Ni_3Al was found to be superplastic at a temperature 450°C below the superplastic temperature in the microcrystalline regime³.

Since the introduction of nanocrystalline materials^{4,5} and the progress in preparation techniques^{6–8} for metals and alloys, many investigators have hoped to obtain superplasticity in pure metals and alloys with nanocrystalline structure at low temperature. This hope is broadly based on the fact that grain-boundary sliding is an important deformation mechanism during superplasticity. Related

to grain-boundary sliding is a grain-size effect that is manifested as a decrease in superplastic temperature or flow stress, or as an increase in the superplastic strain rate, with reduction in grain size¹. For this reason, interest in nanocrystalline materials has been shared by scientists working on superplasticity, because the large volume fraction of grain boundaries in nanostructured materials could lead to low-temperature or high-strain-rate superplasticity, in addition to providing new possibilities of studying grain-boundary-related phenomenon. The work presented here represents a large part of the total experimental data collected to date on superplastic behaviour in such systems.

Tensile samples having a gauge length and width of 1.0 mm, with a nominal thickness of 0.2 mm, were machined by electro-discharge from nickel, 1420-Al and Ni_3Al . The nickel was electrodeposited, and had a purity of greater than 99.5%. The 1420-Al and the Ni_3Al were processed by severe plastic deformation⁹, and had compositions, in weight per cent, of Al-5Mg-2Li-0.1Zr and Ni-8.5Al-7.8Cr-0.6Zr-0.02B, respectively. Specimen size was dictated by

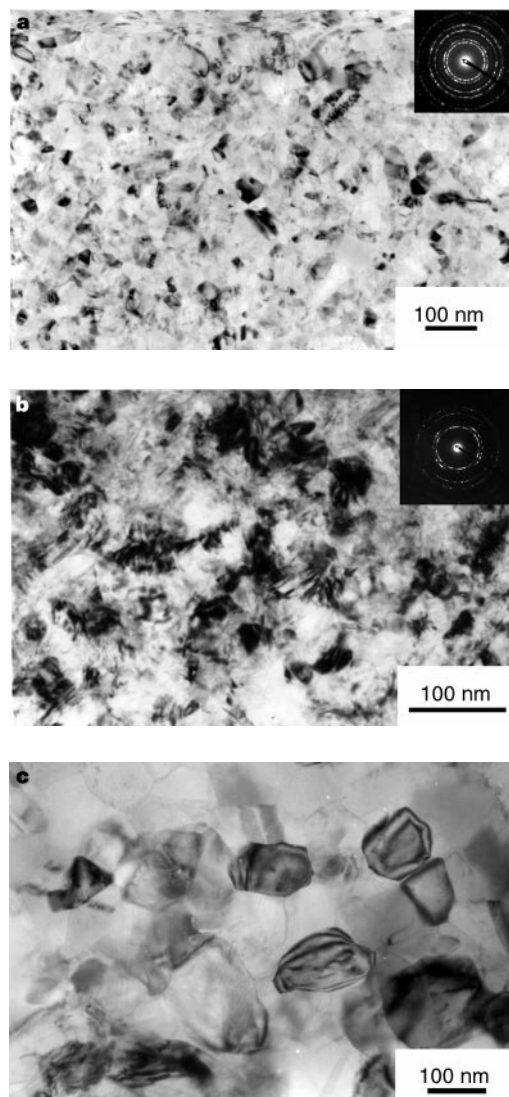


Figure 1 TEM images of the microstructures of the nanocrystalline metals and alloys investigated in this work. Bright-field micrographs show the microstructures of electrodeposited nickel (a), nickel aluminide, Ni_3Al , processed by severe plastic deformation (b), and Ni_3Al after superplastic deformation at 650°C (c). Insets show the ring diffraction patterns, typical of nanostructured materials. Grain growth during deformation was limited in Ni_3Al .

limitations on the size of sample that could be processed by severe plastic deformation; in this process, small disks of 10–15 mm diameter are strained by torsion under high pressure.

As-processed grain sizes for the nickel, 1420-Al and Ni_3Al were 20 nm, 100 nm and 50 nm, respectively. Bright-field transmission electron micrographs of nickel and Ni_3Al are shown in Fig. 1. The tensile samples were deformed in tension at constant strain rates and temperatures. Data on the thermal stability of the nickel are available^{9,10}, and were confirmed in this work using electron microscopy of annealed samples, which showed that rapid grain growth began at 350 °C. Data on the thermal stability of the 1420-Al and Ni_3Al were collected using transmission electron microscopy

(TEM) while heating *in situ*.

Stress–strain curves (Fig. 2) reveal some important features of the mechanical behaviour of these materials. In the case of nickel, a transition occurred above 280 °C; this is indicated by the curve for 350 °C, which shows yielding and strong strain hardening with a large increase in plastic deformation into the superplastic regime of >200% elongation¹. Annealed microcrystalline nickel also exhibits a drop in yield strength and tensile strength within this temperature range¹¹, but the drop was much larger for the nanocrystalline samples and, more importantly, was accompanied by superplastic deformation not observed in larger-grained materials. This transition from low plasticity to superplasticity coincided with the onset of grain growth, and may be explained by a combination of the activation of grain-boundary sliding, increased diffusion, and increased ease of dislocation generation.

The stress–strain curves for nickel, 1420-Al and Ni_3Al demonstrated significant work hardening, and flow stresses were higher than those observed³ for superplastic deformation of microcrystalline states. Microcrystalline Ni_3Al has a yield-stress maximum¹² around 650 °C, referred to in the literature as a “yield stress anomaly”, but this strength maximum is accompanied by a ductility minimum, around 20% total elongation. However, in nanocrystalline Ni_3Al we achieved a tensile ductility of 350% at this temperature. As indicated in Fig. 2, in addition to lower superplastic temperatures 1420-Al also demonstrated high-strain-rate superplasticity, defined by superplastic strain rates greater than 10^{-2} s^{-1} . Such behaviour in a microcrystalline 1420-Al alloy has been reported by Berbon *et al.*¹³. The combination of low-temperature and high-strain-rate superplasticity is significant, and may have technological implications for superplastic forming. Samples before and after deformation are shown in Fig. 3, which demonstrates the large uniform elongation representative of superplasticity.

Based on the heating profile for the tests of nickel at 350 °C, samples were annealed to determine grain size at the start of deformation. The resulting microstructure consisted of a nanocrystalline matrix with isolated 0.3- μm grains. After deformation at 350 °C, the average grain size in the deformed gauge section was 1.3 μm along the tensile axis and 0.64 μm transverse to the tensile axis. Although significant grain growth occurred during deformation, these grain sizes were still much smaller than those which are obtainable using conventional refinement techniques¹⁴. Heating experiments showed that the grain size of 1420-Al and Ni_3Al was unchanged by heating to the respective test temperatures of 250 °C and 650 °C, but that internal strains were reduced, as shown by sharper grain contrast in TEM images. After testing 1420-Al at 250 °C, the grain size in the deformed gauge section was 0.22 μm . Thus, some grain growth occurred during deformation but not to the extent as in electro-deposited nickel. After Ni_3Al was tested at 650 °C, the grain size in the deformed gauge section was 100 nm

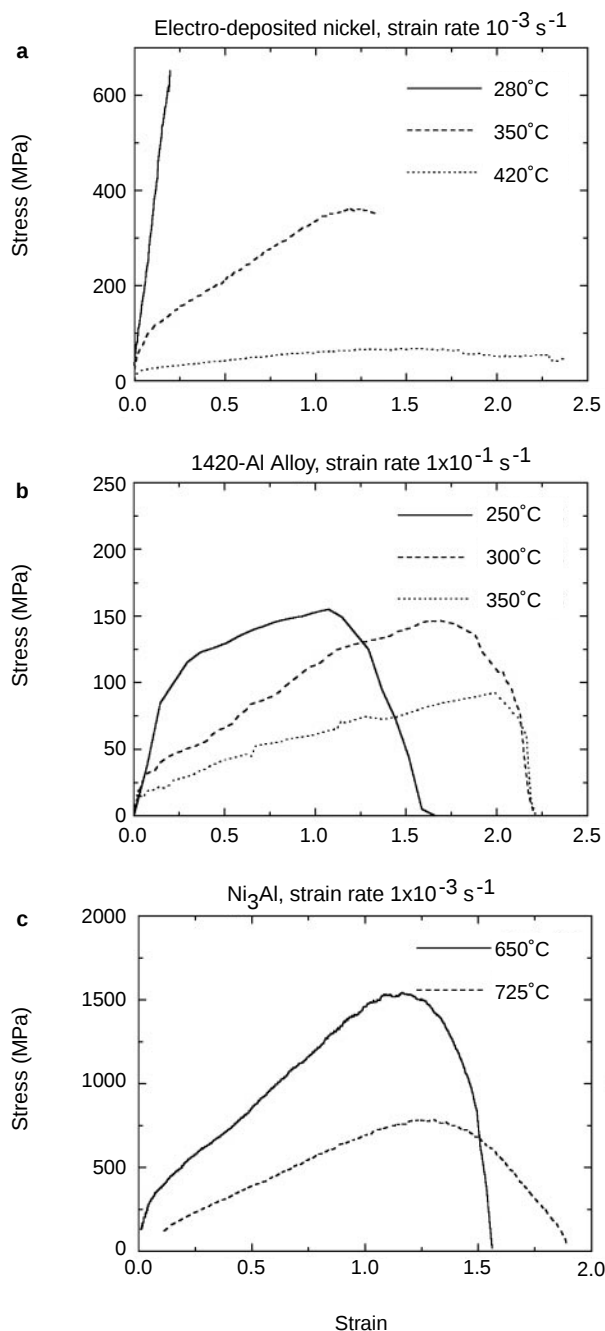


Figure 2 Stress–strain data obtained at constant strain rates and temperatures. Shown are stress–strain curves for electro-deposited nickel (**a**), aluminium alloy 1420-Al processed by severe plastic deformation (**b**), and Ni_3Al processed by severe plastic deformation (**c**). Note the transition from low plasticity to superplasticity in **a** between 280 °C and 350 °C, the high strain rates in **b**, and the very high flow stresses in **c**.

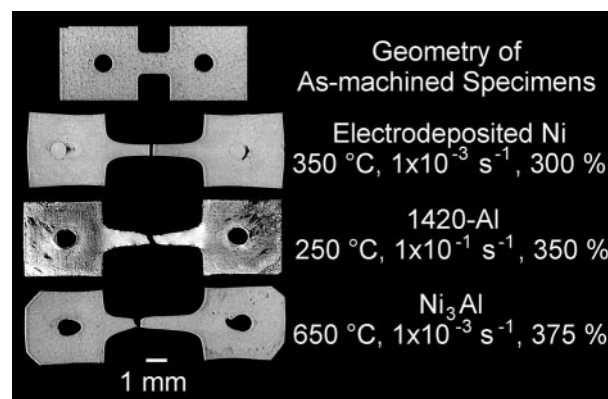


Figure 3 Miniature tensile specimens, shown in the as-machined geometry, and after deformation.

(Fig. 1c). Therefore, Ni₃Al remained nanocrystalline throughout the deformation. This observation has significance regarding the deformation mechanism. The common explanation of strain hardening during superplasticity at constant strain rate and temperature is that grain growth increases the flow stress through grain-size sensitivity, where stress is proportional to grain size. But at 650 °C, the flow stress for nanocrystalline Ni₃Al increased six-fold between the yield stress and the peak stress, while the grain size increased only two-fold. This suggests that mechanistic details of superplasticity in nanocrystalline materials are fundamentally different from those in microcrystalline materials. A detailed microstructural investigation is required to explain the high strain hardening observed during superplasticity.

From the stress-strain curves, it is clear that the maximum elongation for each material was found at higher temperatures than those discussed above. The reason for focusing on the lower temperatures is to investigate the effect of nanostructure, and to probe the low-temperature limits of superplasticity. At higher temperatures, grain growth shifts the microstructure into the sub-microcrystalline range. Furthermore, the practical advantages of superplastic deformation are not necessarily lost by operating away from the ductility maximum, as deformation in the range of a few hundred per cent is all that is required for many commercial forming operations.

Our experimental results allow us to make some observations about superplasticity in nanocrystalline metals and alloys. The single most significant observation is the large reduction in superplastic temperatures. In the case of nickel, an extremely low normalized temperature was attained. But even at very low temperatures, grain growth during deformation of nanostructured nickel resulted in a grain size larger than the commonly used definition for nanostructured materials of 100 nm or less. The primary reason for such growth is a large driving force resulting from grain-boundary energy. These results appear to eliminate the hope of obtaining superplasticity in pure metals having a grain size of 100 nm or less, because the reduction of superplastic temperature is offset by a reduction in the grain-growth temperature. However, secondary factors such as internal strain energy may influence the grain-growth temperature through contributions to the driving force, and these effects have not yet been evaluated.

Even if secondary contributions to grain growth could be eliminated, grain-growth and superplastic temperatures cannot be treated independently because both are thermally activated processes linked closely through solid-state diffusion. Therefore, in the limiting case of nanostructured pure metals, grain growth will occur if sufficient thermal activation is present for superplastic deformation. This competition between grain growth and superplasticity was established early in the development of the field, but has not hitherto been investigated at such small starting grain size.

Grain growth is commonly controlled with multiple-phase alloys, where second phases are present as particles or individual grains. Particles inhibit grain growth by pinning the boundaries, while two-phase systems require diffusion through neighbouring grains of different composition to accommodate growth. These effects can be seen by comparing the grain growth during deformation between nanocrystalline nickel and 1420-Al (data not shown), where the increase in grain size of nickel was more than that of 1420-Al even though nickel was at a lower normalized temperature. In the case of ordered intermetallics, such as Ni₃Al, grain growth is inhibited by the kinetic barrier of preferred atomic pairing between the species present, and the grain-size stability of the Ni₃Al is attributed to this effect.

Investigation of superplasticity in nanostructured materials is still in its infancy, and many systems exist that hold promise for superplastic behaviour while retaining a grain size of 100 nm or less. The small number of experimental results on superplasticity obtained from nanostructured materials have upheld predictions of low-temperature deformation, while at the same time the difficulty

of maintaining a nanocrystalline grain size in pure nickel has emphasized the interaction of thermally activated processes in these materials. However, the high flow stresses and significant work hardening observed in Ni₃Al and 1420-Al have suggested that details of superplasticity are fundamentally different in nanostructured materials. □

Received 3 November 1998; accepted 19 February 1999.

1. Mukherjee, A. K. in *Materials Science and Technology* Vol. 6, *Plastic Deformation and Fracture of Materials* (ed. Mughrabi, H.) Ch. 9 (VCH, New York, 1993).
2. Floren, S. Superplasticity in pure nickel. *Scripta Metall.* **1**, 19–23 (1967).
3. Mukhopadhyay, J., Kashner, G. & Mukherjee, A. K. Superplasticity in boron doped Ni₃Al alloy. *Scripta Metall. Mater.* **24**, 857–862 (1990).
4. Gleiter, H. Nanocrystalline materials. *Prog. Mater. Sci.* **33**, 223–315 (1989).
5. Suryanarayana, C. Nanocrystalline materials. *Int. Mater. Rev.* **40**, 41–64 (1995).
6. Valiev, R. Z. Ultrafine-grained materials produced by severe plastic deformation. *Ann. Chim.* **21**, 6–7, 369–378 (1996).
7. Birringer, R. & Gleiter, H. in *Encyclopedia of Material Science and Engineering: Supplement 1* (eds Cahn, R. W. & Beaver, M. B.) 339–349 (Pergamon, Oxford, 1988).
8. El-Sherik, A. M. & Erb, U. Synthesis of bulk nanocrystalline nickel by pulsed electrodeposition. *J. Mater. Sci.* **30**, 5743–5749 (1995).
9. Wang, N., Wang, Z., Aust, K. T. & Erb, U. Isochronic analysis of nanocrystalline nickel electrodeposits upon annealing. *Acta Mater.* **45**, 1655–1669 (1997).
10. Natter, H., Schmeltzer, M. & Hempelmann, R. Nanocrystalline nickel and nickel copper alloys: Synthesis, characterisation, and thermal stability. *J. Mater. Res.* **13**, 1186–1197 (1998).
11. Everhart, J. L. *Engineering Properties of Nickel and Nickel Alloys* 20 (Plenum, New York, 1971).
12. Liu, V. T. & Sikka, V. Nickel aluminides for structural use. *J. Metals* **38**, 19–21 (1987).
13. Berbon, P. B. *et al.* Fabrication of bulk ultrafine-grained materials through intense plastic straining. *Metall. Mater. Trans. A* **29**, 2237–2243 (1998).
14. Kaybyshev, O. A. & Mareklov, A. A. Structural changes during superplastic deformation of nickel and chromium. *Fiz. Metal. Metalloved.* **41**, 190–196 (1976).

Acknowledgements. This work was supported by the US National Science Foundation and the Civil Research and Development Foundation.

Correspondence and requests for materials should be addressed to A.K.M. (e-mail: akmukherjee@ucdavis.edu).

Stochastic sensing of organic analytes by a pore-forming protein containing a molecular adapter

Li-Qun Gu*, Orit Braha*, Sean Conlan*, Stephen Cheley* & Hagan Bayley*†

* Department of Medical Biochemistry & Genetics, Texas A&M University Health Science Center, 440 Reynolds Medical Building, College Station, Texas 77843-1114, USA

† Department of Chemistry, Texas A&M University, College Station, Texas 77843-3255, USA

The detection of organic molecules is important in many areas, including medicine, environmental monitoring and defence^{1–5}. Stochastic sensing is an approach that relies on the observation of individual binding events between analyte molecules and a single receptor⁶. Engineered transmembrane protein pores are promising sensor elements for stochastic detection⁶, and in their simplest manifestation they produce a fluctuating binary ('on/off') response in the transmembrane electrical current. The frequency of occurrence of the fluctuations reveals the concentration of the analyte, and its identity can be deduced from the characteristic magnitude and/or duration of the fluctuations. Genetically engineered versions of the bacterial pore-forming protein α -haemolysin have been used to identify and quantify divalent metal ions in solution⁶. But it is not immediately obvious how versatile binding sites for organic ligands might be obtained by engineering of the pore structure. Here we show that stochastic sensing of organic molecules can be procured from α -haemolysin by equipping the channel with an internal, non-covalently bound molecular 'adapter' which mediates channel blocking by the analyte. We use cyclodextrins as the adapters because these fit

comfortably inside the pore and present a hydrophobic cavity suitable for binding a variety of organic analytes. Moreover, a single sensing element of this sort can be used to analyse a mixture of organic molecules with different binding characteristics. We envisage the use of other adapters, so that the pore could be 'programmed' for a range of sensing functions.

α -Haemolysin is an exotoxin secreted by the bacterium *Staphylococcus aureus*⁷. The monomeric 293-amino-acid polypeptide can self-assemble on lipid bilayers to form a heptameric pore. Transported molecules move through a 100 Å-long channel centred on the molecular 7-fold axis⁸. The opening of the channel on the *cis* side of the bilayer (the side of assembly) is ~70 Å above the membrane surface and 29 Å in diameter. The channel widens into a roughly spherical vestibule of ~42 Å in diameter. About 20 Å above the membrane plane, the vestibule narrows and becomes a 14-stranded β -barrel, 52 Å in length, which continues through the membrane with an average internal diameter of about 20 Å. The *trans* entrance to the channel lies close to the bilayer surface. Roughly globular molecules of relative molecular mass up to ~2,000 (2K; refs 9, 10), or larger, elongated polymers such as single-stranded nucleic acids¹¹, can pass through the α -haemolysin pore.

We found that α -, β - and γ -cyclodextrins at micromolar concentrations enter the α -haemolysin (α HL) channel and produce

reversible partial blocks of the ionic current. β -Cyclodextrin (β CD; $M_r = 1.135K$) was investigated further. The block was established when β -cyclodextrin was added from the *trans* side of a planar bilayer (Fig. 1a, b), but not from the *cis* side. The kinetics of blocking were consistent with a simple scheme in which there is a single binding site for β -cyclodextrin inside the lumen of the channel, for which $k_{on}^C = 5.46 \times 10^4 M^{-1} s^{-1}$, $k_{off}^C = 1.15 \times 10^2 s^{-1}$, and the formation constant $K_f^C = 4.75 \times 10^2 M^{-1}$ in 1 M NaCl, 10 mM sodium phosphate and 5 μ M EDTA at pH 3.0 (*trans*). The conductance of α -haemolysin was 750 pS (s.d. = 21, $n = 35$) in the absence of β -cyclodextrin, and 272 pS (s.d. = 9, $n = 35$) in its presence. Molecular modelling revealed that β -cyclodextrin could fit into the lumen of the α -haemolysin channel, with its molecular 7-fold axis coincident with the 7-fold axis of the heptameric pore. On entering from the *trans* side, β -cyclodextrin might be retained by the ring of seven leucine residues (Leu 135) that are found about half-way through the transmembrane domain or, after squeezing past this restriction, it might be stopped by a second restriction near the *cis* end of the β -barrel which comprises (from the *trans* side) the amino-acid residues Met 113, Lys 147 and Glu 111. We distinguished between these two possibilities by site-directed mutagenesis. The kinetics of block by β -cyclodextrin in the mutant L135N (carrying an asparagine instead of a leucine residue at

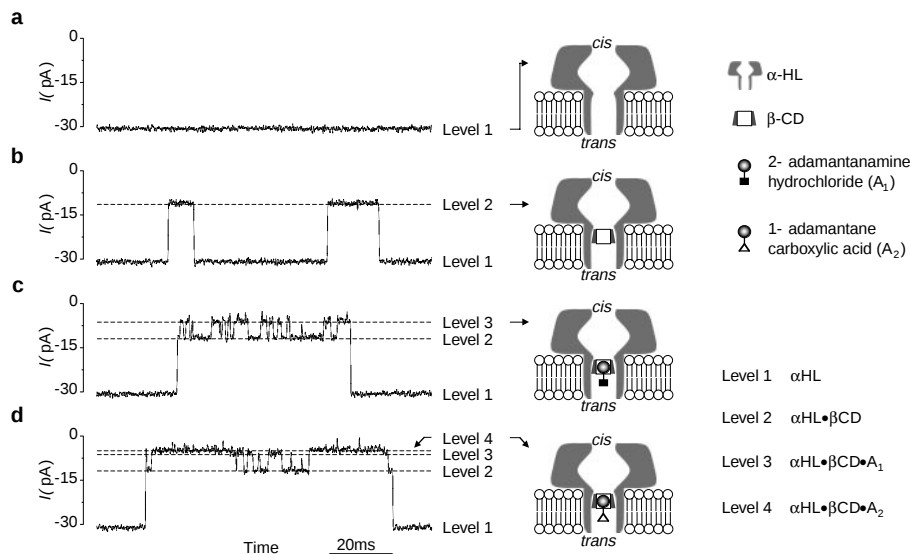


Figure 1 Bilayer recordings showing the interaction of a single α -haemolysin (α HL) pore with β -cyclodextrin (β CD) and the model analytes 2-adamantanamine (A_1) and 1-adamantanecarboxylic acid (A_2). All traces were recorded at -40 mV (*cis* at ground). α HL was added to the *cis* chamber and β CD and the adamantane derivatives to the *trans* chamber. **a**, Single α -HL pore continuously open, -31.5 pA (level 1); **b**, β CD (20 μ M, *trans*) produces transient partial blockades of the

channel, -11.5 pA (level 2); **c**, 2-adamantanamine (80 μ M, *trans*) does not affect the fully open channel (level 1), but produces an additional block of α HL $\cdot$ β CD, -5.7 pA (level 3); **d**, 1-adamantanecarboxylic acid (20 μ M, *trans*) produces additional blockades, -4.7 pA (level 4), of longer duration than those produced by 2-adamantanamine (level 3).

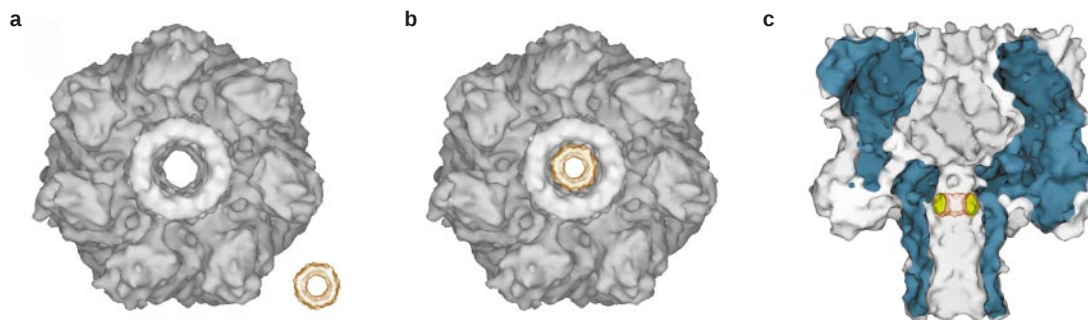


Figure 2 Molecular graphics representation of the interaction between α HL and β CD. **a**, View into the channel from the *trans* side, β CD removed. **b**, View into the channel from the *trans* side with β CD (gold) bound. **c**, Sagittal section through the heptameric α -haemolysin pore showing β -cyclodextrin (yellow and gold) lodged

in the lumen of the transmembrane channel. On the basis of mutagenesis data, β CD is shown placed against a constriction in the transmembrane β -barrel formed by the ring of seven Met 113 side chains.

position 135) were similar to those of wild-type (WT) α -haemolysin (pH 7.5 (*trans*): WT, $k_{\text{off}}^C = 1.08 \times 10^3 \text{ s}^{-1}$; L135N, $k_{\text{off}}^C = 4.7 \times 10^2 \text{ s}^{-1}$). By contrast, the blocking events were greatly prolonged for both M113N and the double mutant M113N/L135N (pH 7.5 (*trans*): M113N, $k_{\text{off}}^C \approx 3.7 \times 10^{-2} \text{ s}^{-1}$; M113N/L135N, $k_{\text{off}}^C \approx 0.13 \text{ s}^{-1}$). These results suggest that β -cyclodextrin binds near Met 113 in wild-type α -haemolysin (Fig. 2) and that when this residue is mutated the binding affinity is increased at the second restriction, which is now formed by Lys 147 and Glu 111.

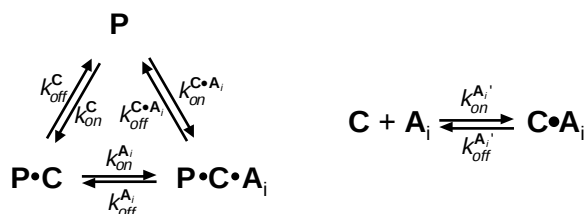


Figure 3 A kinetic scheme for the interactions between α -haemolysin (α HL), β -cyclodextrin (β CD) and guests. P, α HL pore; C, the adapter β CD; A_i , guests (analytes); $C \cdot A_i$, $P \cdot C$, $P \cdot C \cdot A_i$, the various non-covalent complexes of P, C and A_i .

Because the block of wild-type α -haemolysin by β -cyclodextrin was partial (64%) and because cyclodextrins act as hosts for a variety of guest molecules, we investigated whether guests could further reduce single-channel currents. If they could, then the system would constitute a stochastic sensor for organic molecules. A large number of readily available adamantane derivatives bind to cyclodextrins¹² and were therefore chosen as model analytes. Strikingly, 2-adamantanamine (A_1 , $80 \mu\text{M}$ *trans*) reduced the conductance of the partially blocked channel to 126.5 pS (s.d. = 0.5; $n = 7$) with a residence time (τ_{A1}) of 2.54 ms (s.d. = 0.21), but had no effect on the completely open channel (Fig. 1c). A second guest, 1-adamantanecarboxylic acid (A_2 , $20 \mu\text{M}$ *trans*), also reduced the conductance of the partially blocked channel, this time to 112.2 pS (s.d. = 3.2, $n = 7$) with a residence time (τ_{A2}) of 14.0 ms (s.d. = 0.8). During each β -cyclodextrin blocking event, guest residence times were independent of guest concentration and the rates of guest association were linearly dependent on concentration, suggestive of a bimolecular interaction. Events due to each guest in a mixture could be distinguished, indicating that they compete for a single binding site in the α -haemolysin $\cdot$ β -cyclodextrin complex (Fig. 1d). A simple kinetic scheme¹³ is sufficient to explain the observed interactions of analytes in simple solutions or as mixtures

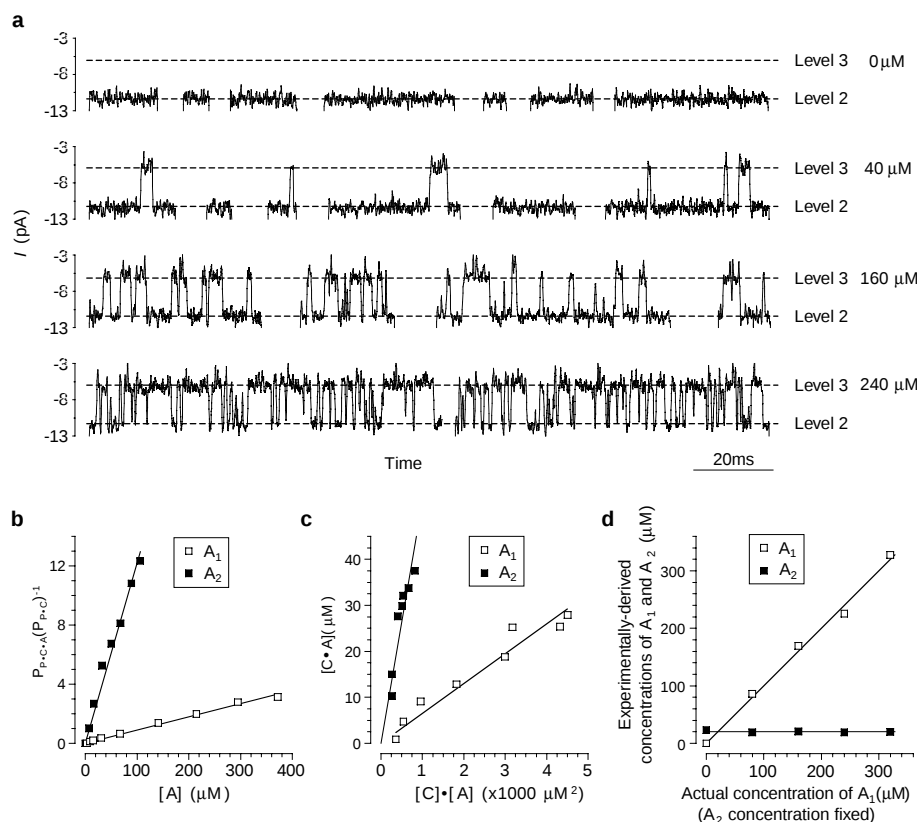


Figure 4 Response of α HL $\cdot$ β CD at different analyte concentrations. **a**, α HL $\cdot$ β CD (level 2) to α HL $\cdot$ β CD $\cdot$ A_1 (level 3) transitions at various 2-adamantanamine (A_1) concentrations. Representative segments of traces are shown depicting entire β CD occupancy events at 0, 40, 160 and 240 μM 2-adamantanamine. The conditions were as for Fig. 1, but with 40 μM β CD. **b**, Plots of $P_{P \cdot C \cdot A_1}/P_{P \cdot C}$ versus $[A]$ for 2-adamantanamine (A_1) and 1-adamantanecarboxylic acid (A_2) using data from **a**. The slope yields the formation constant K_f^A for α HL $\cdot$ β CD $\cdot$ A from analyte A and α HL $\cdot$ β CD. For a single analyte, $P_{P \cdot C \cdot A}/P_{P \cdot C} = K_f^A [A]$, where $P_{P \cdot C \cdot A}$ and $P_{P \cdot C}$ are the experimentally determined probabilities of occurrence of the states of the α HL pore with both β CD and analyte bound or with only β CD bound, K_f^A is the equilibrium formation constant for α HL $\cdot$ β CD $\cdot$ A from analyte A and α HL $\cdot$ β CD, and $[A]$ is the concentration of free analyte. $[A]$ can be determined from the experimental values of $P_{P \cdot C}$ and P_P (P_P is the probability of occurrence of the

unoccupied α HL pore), $[A]_0$ and $[C]_0$ (the total concentrations of analyte and β CD) and K_f^C (the equilibrium formation constant for the α HL $\cdot$ β CD complex), which can be measured separately (see Supplementary Information). **c**, Plots of $[C \cdot A]$ versus $[C][A]$ for 2-adamantanamine (A_1) and 1-adamantanecarboxylic acid (A_2) using data from **a**. The slope yields the formation constant K_f^A for β CD $\cdot$ A from analyte and β CD in solution. $[C \cdot A] = K_f^A [C][A]$, where values for $[C \cdot A]$, $[C]$ and $[A]$, the concentrations of the β CD $\cdot$ A complex, free β CD and free analyte, respectively, can be obtained from experimental or known values of $P_{P \cdot C}$, P_A , $[A]_0$, $[C]_0$ and K_f^C (see Supplementary Information). **d**, Analysis of currents from binary solutions of analytes. The experimentally derived concentrations of 2-adamantanamine (A_1) and 1-adamantanecarboxylic acid (A_2), determined by applying equation (1) to current amplitude histograms, are plotted against the actual concentration of A_1 . Conditions are as for Fig. 1, but with 40 μM β CD.

(Fig. 3; details in Supplementary Information). In the case of β -cyclodextrin and the adamantane derivatives described here, the residence time of the cyclodextrin in the lumen of the pore is long compared with the residence time of the analyte in the cyclodextrin. Therefore, β -cyclodextrin acts as a non-covalent molecular adapter. When the residence time of the analyte within the cyclodextrin host is relatively long, the host acts instead as a carrier.

The signals can be used not only to identify analytes but also to quantify them. As expected, the frequency of occurrence of α HL- β CD occupancy events by an analyte increases with analyte concentration (Fig. 4a). Interestingly, the residence time of β CD-2-adamantanamine on α -haemolysin is greater than β -cyclodextrin itself (Fig. 4a), which is true for all β CD-A observed in this study. From the slope of a linear plot (Fig. 4b), K_f^A values for 1-adamantanecarboxylic acid and 2-adamantanamine (at pH 3.0) were found to be, respectively, $1.35 \pm 0.19 \times 10^5 \text{ M}^{-1}$ and $1.03 \pm 0.15 \times 10^4 \text{ M}^{-1}$ (mean $\pm$ s.d., $n = 7$), corresponding to ΔG values of $-7.0 \text{ kcal mol}^{-1}$ and $-5.5 \text{ kcal mol}^{-1}$. Values of K_f^A , the equilibrium formation constant for the analyte (A) with β -cyclodextrin in solution, can be obtained from a second plot (Fig. 4c). K_f^A values at pH 3.0 for 1-adamantanecarboxylic acid and 2-adamantanamine were, respectively, $5.76 \pm 1.50 \times 10^4 \text{ M}^{-1}$ and $9.84 \pm 2.19 \times 10^3 \text{ M}^{-1}$ (mean $\pm$ s.d., $n = 7$), corresponding to ΔG values of $-6.5 \text{ kcal mol}^{-1}$ and $-5.5 \text{ kcal mol}^{-1}$, which are close to the values found when β -cyclodextrin is bound to the α -haemolysin pore. Literature values for ΔG in solution are in the same range as those we have determined: 1-adamantanecarboxylic acid, $-7.5 \text{ kcal mol}^{-1}$ (pH 4.05)¹⁴; 2-adamantanamine, $-5.3 \text{ kcal mol}^{-1}$ (pH 2.5)¹². In a working sensor, the total concentration $[A_i]_0$ of analyte A_i of known K_f^{Ai} and $K_f^{Ai'}$ would be determined from:

$$[A_i]_0 = (1/K_f^{Ai})(1/P_{P-C} + K_f^{Ai'}/(P_P K_f^C))P_{P-C:Ai} \quad (1)$$

where $P_{P-C:A}$ and P_{P-C} are the experimentally determined probabilities of occurrence of the states of the α -haemolysin pore with both β -cyclodextrin and analyte bound, or with only β -cyclodextrin

bound, and P_P is the probability of occurrence of the unoccupied α -haemolysin pore (see Supplementary Information for details).

Another important attribute of stochastic sensing is that two or more analytes, only one of which can occupy the receptor at a given moment, can be identified and quantified 'simultaneously' by a single sensor element. In a mixture, signals from different analytes are recognized by their characteristic extents of channel block and residence times¹⁵ (for example, see Table 1 of Supplementary Information for values for seven adamantanes). To illustrate this with α -haemolysin and the β -cyclodextrin adapter, we did an experiment in which 1-adamantanecarboxylic acid (A_2) was kept constant at $20 \mu\text{M}$, while 2-adamantanamine (A_1) was varied. Current amplitude histograms, which revealed P_B , P_{P-C} and $P_{P-C:Ai}$ and equation (1) were used to generate the total concentrations of the two analytes, $[A_1]_0$ and $[A_2]_0$. The values obtained were in close agreement with the actual concentrations in the mixtures (Fig. 4d).

The principle of stochastic sensing with adapter molecules can be expanded to molecules of biological interest. For example, many pharmaceuticals interact with cyclodextrins¹² and many of those we have tested give useful signals in our system. For instance, the tricyclic antidepressant imipramine can be distinguished by using residence-time histograms from the antihistaminic promethazine, which has a similar structure (Fig. 5a–c). The use of alternative cyclodextrins, as illustrated here by the interaction of imipramine with γ -cyclodextrin (Fig. 5d), demonstrates that the system is programmable. Stochastic sensing with adapter molecules is therefore a versatile approach to analyte identification and quantification. As in olfaction, in which carrier molecules (olfactory binding proteins) deliver odorants to membrane-bound receptors¹⁶, we could vary both the adapter/carrier and the receptor: for example, the range of detectable analytes will be extended by using naturally occurring and chemically modified cyclodextrins¹² as adapters; synthetic adapter/carriers could also be considered^{17–19}. Genetically engineered α -haemolysin pores²⁰, or pores other than α -haemolysin^{21,22} could be used to accommodate the alternative

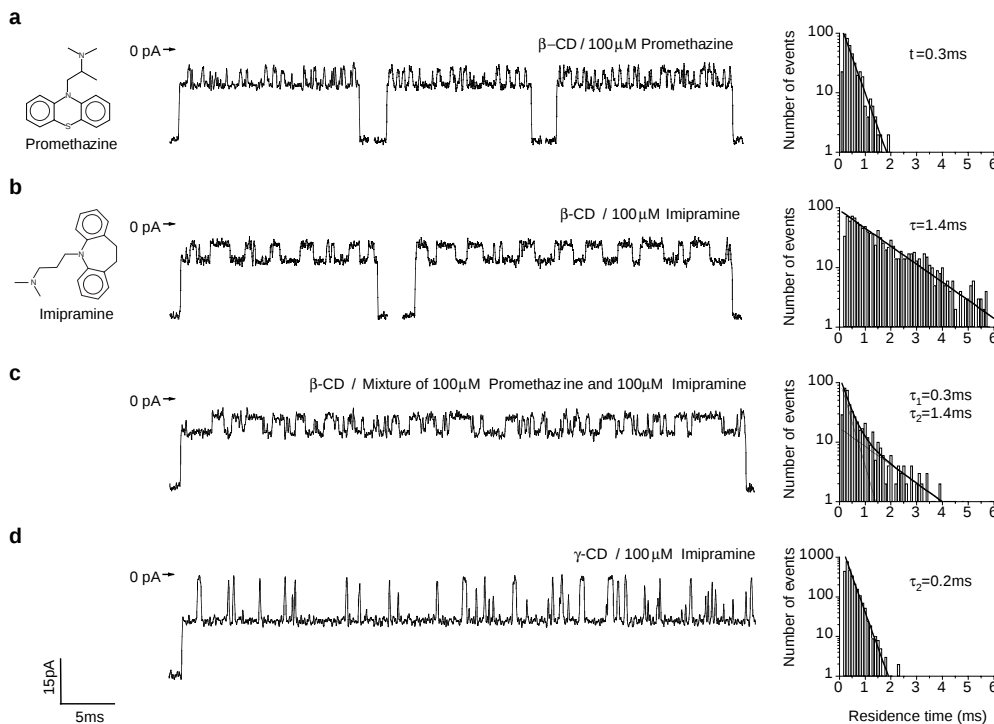


Figure 5 Analysis of drug molecules by stochastic sensing with α HL and a β CD or γ CD adapter. **a**, Signal from $100 \mu\text{M}$ promethazine in the presence of $40 \mu\text{M}$ β CD. Many events with a mean duration of 0.3 ms are observed. **b**, Signal from $100 \mu\text{M}$ imipramine in the presence of $40 \mu\text{M}$ β CD. The extent of channel block is similar

to that of promethazine, but a characteristic signature is provided by the longer residence time of 1.4 ms . **c**, Mixture of $100 \mu\text{M}$ promethazine and $100 \mu\text{M}$ imipramine in the presence of $40 \mu\text{M}$ β CD. **d**, Signal from $100 \mu\text{M}$ imipramine in the presence of $20 \mu\text{M}$ γ CD.

adapters, which might also be covalently attached to the pores. Single-molecule fluorescence²³ or force-detection methods²⁴ might be feasible for read-out. For applications in the field, rugged stochastic sensors must be developed, as was recently achieved for a sensor based on macroscopic channel currents²⁵. □

Methods

Pore formation. Heptameric wild-type α -haemolysin formed by treating the monomer, purified from *S. aureus*, with deoxycholate^{26,27} was isolated from SDS-polyacrylamide gels as described⁶.

Planar bilayer recordings. A bilayer of 1,2-diphytanoylphosphatidylcholine (Avanti Polar Lipids) was formed on a 100- to 150- μ m orifice in a 25- μ m-thick Teflon film (Goodfellow) separating two compartments (2 ml each) of a planar bilayer apparatus²⁸. Solutions in the compartments contained 1 M NaCl, 5 μ M EDTA, 10 mM sodium phosphate, at pH 7.5 in the *cis* chamber and pH 3.0 in the *trans* chamber. Neutral-pH solutions in the *cis* chamber prevent closure of α -haemolysin channels; low-pH solutions in the *trans* chamber increase the residence time of cyclodextrins. Heptameric α -haemolysin (0.5 to 1 μ l at 50 to 200 ng ml⁻¹) was added to the *cis* compartment, which was stirred until a single channel inserted into the bilayer. Currents were recorded at a holding potential of -40 mV (*cis* at ground) by using a patch-clamp amplifier (Axopatch 200B, Axon Instruments). Currents were low-pass filtered with a built-in 4-pole Bessel filter at 5 kHz and sampled at 20 kHz by computer with a Digidata 1200 A/D converter (Axon Instruments). Cyclodextrins (Aldrich) and analytes (Aldrich) were added to the *trans* chamber.

Data analysis. Current amplitude and residence (dwell) time histograms were constructed using pClamp 6.0 software (Axon Instruments). Probabilities, P_B , P_{P-C} and P_{P-C_A} , were obtained from the amplitude histograms after fitting the peaks to gaussian functions. Mean residence times (τ values) were obtained from residence time histograms by fitting the distributions to exponential functions. Measurements are given as the mean $\pm$ s.d.

Molecular graphics. The structures of α -haemolysin were generated from the coordinates⁸ by using SPOCK 6.3 software²⁹. The β -cyclodextrin structure was generated with Insight II 97.0. The two structures were scaled and superimposed in Adobe Photoshop 4.0.

Received 6 January; accepted 3 March 1999.

- Dickinson, T. A., White, J., Kauer, J. S. & White, D. R. A chemical-detecting system based on a cross-reference optical sensor array. *Nature* **382**, 687–700 (1996).
- Loneragan, M. C. et al. Array-based vapor sensing using chemically sensitive, carbon black-polymer resistors. *Chem. Mat.* **8**, 2298–2312 (1996).
- Hellinga, H. W. & Marvin, J. S. Protein engineering and the development of generic biosensors. *Trends Biotechnol.* **16**, 183–189 (1998).
- Czarnik, A. W. A sense for landmines. *Nature* **394**, 417–418 (1998).
- Crooks, R. M. & Ricco, A. J. Special issue on chemical sensors. *Acc. Chem. Res.* **31**, 199–324 (1998).
- Braha, O. et al. Designed protein pores as components for biosensors. *Chem. Biol.* **4**, 497–505 (1997).
- Gouaux, E. α -Haemolysin from *Staphylococcus aureus*: an archetype of β -barrel, channel-forming toxins. *J. Struct. Biol.* **121**, 110–122 (1998).
- Song, L. et al. Structure of staphylococcal α -hemolysin, a heptameric transmembrane pore. *Science* **274**, 1859–1865 (1996).
- Krasnik, O. V., Sabirov, R. Z., Ternovsky, V. I., Merzliak, P. G. & Muratkhodjaev, J. N. A simple method for the determination of the pore radius of ions channels in planar lipid bilayer membranes. *FEMS Microbiol. Immunol.* **105**, 93–100 (1992).
- Bezrukov, S. M., Vodyanov, I., Brutyan, R. A. & Kasianowicz, J. J. Dynamics and free energy of polymer partitioning into a nanoscale pore. *Macromolecules* **29**, 8517–8522 (1996).
- Kasianowicz, J. J., Brandin, E., Branton, D. & Deamer, D. W. Characterization of individual polynucleotide molecules using a membrane channel. *Proc. Natl Acad. Sci. USA* **93**, 13770–13773 (1996).
- Rekharsky, M. V. & Inoue, Y. Complexation thermodynamics of cyclodextrins. *Chem. Rev.* **98**, 1875–1917 (1998).
- Colquhoun, D. & Hawkes, A. G. On the stochastic properties of bursts of single ion channel openings and of clusters of bursts. *Phil. Trans. R. Soc. Lond. B* **300**, 1–59 (1982).
- Inoue, Y. et al. Thermodynamics of molecular recognition by cyclodextrins. 1. Calorimetric titration of inclusion complexation of naphthalenesulfonates with α -, β -, and γ -cyclodextrins: enthalpy–entropy compensation. *J. Am. Chem. Soc.* **115**, 475–481 (1993).
- Lucchesi, K., Ravindran, A., Young, H. & Moczydlowski, E. Analysis of the blocking activity of charybdotoxin homologs and iodinated derivatives against Ca²⁺-activated K⁺ channels. *J. Membr. Biol.* **109**, 269–281 (1989).
- Bianchet, M. A. et al. The three-dimensional structure of bovine odorant binding protein and its mechanism of odor recognition. *Nature Struct. Biol.* **3**, 934–939 (1996).
- Chen, H., Weiner, W. S. & Hamilton, A. D. Recognition of neutral species with synthetic receptors. *Curr. Opin. Chem. Biol.* **1**, 458–466 (1997).
- Arduini, A., Casnati, A., Pochini, A. & Ungaro, R. Recognition of cationic species with synthetic receptors. *Curr. Opin. Chem. Biol.* **1**, 467–474 (1997).
- Beer, P. D. & Schmitt, P. Molecular recognition of anions by synthetic receptors. *Curr. Opin. Chem. Biol.* **1**, 475–482 (1997).
- Bayley, H. Building doors into cells. *Sci. Am.* **277**, (Sept.) 62–67 (1997).
- Hartgerink, J. D., Clark, T. D. & Ghadiri, M. R. Peptide nanotubes and beyond. *Chem. Eur. J.* **4**, 1367–1372 (1998).

- Schmid, B., Maveyraud, L., Kromer, M. & Schulz, G. E. Porin mutants with new channel properties. *Protein Sci.* **7**, 1603–1611 (1998).
- Lu, H. P., Xun, L. & Xie, X. S. Single-molecule enzymatic dynamics. *Science* **282**, 1877–1882 (1998).
- Oberhauser, A. F., Marszalek, P. E., Erickson, H. P. & Fernandez, J. M. The molecular elasticity of the extracellular matrix protein tenascin. *Nature* **393**, 181–185 (1998).
- Cornell, B. A. et al. A biosensor that uses ion-channel switches. *Nature* **387**, 580–583 (1997).
- Bhakti, S., Füssle, R. & Tranum-Jensen, J. Staphylococcal α -toxin: oligomerization of hydrophilic monomers to form amphiphilic hexamers induced through contact with deoxycholate micelles. *Proc. Natl Acad. Sci. USA* **78**, 5475–5479 (1981).
- Walker, B. J., Krishnaswamy, M., Zorn, L., Kasianowicz, J. J. & Bayley, H. Functional expression of the α -hemolysin of *Staphylococcus aureus* in intact *Escherichia coli* and in cell lysates. *J. Biol. Chem.* **267**, 10902–10909 (1992).
- Montal, M. & Mueller, P. Formation of bimolecular membranes from lipid monolayers and study of their electrical properties. *Proc. Natl Acad. Sci. USA* **69**, 3561–3566 (1972).
- Christopher, J. A. SPOCK: The Structural Properties Observation and Calculation Kit (Program Manual) (Center for Macromolecular Design, Texas A&M Univ., College Station, 1998).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements. This work was supported by DARPA, DOE and ONR. We thank E. Gouaux for comments on the manuscript.

Correspondence and requests for materials should be addressed to O.B. (e-mail: obraha@medicine.tamu.edu) or H.B. (e-mail: bayley@tamu.edu).

Present and future trends in the atmospheric burden of ozone-depleting halogens

S. A. Montzka*, J. H. Butler*, J. W. Elkins*, T. M. Thompson*, A. D. Clarke†‡ & L. T. Lock†

* National Oceanic and Atmospheric Administration, Climate Monitoring and Diagnostics Laboratory, 325 Broadway, Boulder, Colorado 80303, USA

† Cooperative Institute for Research in Environmental Science, University of Colorado, Boulder, Colorado 80309, USA

The burden of ozone-depleting chemicals in the lower atmosphere has been decreasing since 1994 as a result of the Montreal Protocol^{1–3}. Here we show how individual chemicals have influenced this decline, in order to estimate how the burden could change in the near future. Our measurements of atmospheric concentrations of the persistent, anthropogenic chemicals that account for most ozone-depleting halogens in today's stratosphere show that the decline stems predominantly from the decrease in the atmospheric load of trichloroethane (CH₃CCl₃), a previously common cleaning solvent. The influence of this chemical on the decline has now peaked, however, and will become much smaller over the next five to ten years. As this influence lessens, a decrease in the burden of ozone-depleting halogen will be sustained only if emissions of other halocarbons fall. Although emissions of most gases regulated by the Montreal Protocol have decreased substantially over the past ten years (refs 4–11), emissions of the potent ozone-depleting gas CBrClF₂ (halon-1211) have remained fairly constant during this period^{12,29}, despite stringent limits on production in developed countries since 1994. The consequent atmospheric accumulation of this halon is retarding the decline of ozone-depleting halogens in the atmosphere more than any other persistent gas.

Measurements of chlorinated and brominated trace gases at remote locations on the Earth's surface provide constraints on the burden of ozone-depleting chemicals in the future stratosphere^{1,4,13,14}. Here we relate tropospheric halocarbon burdens to the future stratospheric abundance of ozone-depleting halogens by considering halocarbon destruction rates in the stratosphere and by recognizing that bromine is 50 times more effective than chlorine in ozone-depleting reactions⁴ (Table 1). The aggregate halocarbon burden weighted in this way, which is called "effective equivalent

‡ Present address: National Oceanic and Atmospheric Administration, Climate Monitoring and Diagnostics Laboratory, South Pole, Antarctica.

chlorine" (EECl)^{1,4,14}, declined in 1996–97 at rates similar to those reported for mid-1995¹ (Figs 1 and 2). By the end of 1997, a total decrease of about 90 parts per 10¹² (parts per trillion, p.p.t.) EECl from the maximum observed in 1993–94 is estimated based on NOAA/CMDL measurements of anthropogenic halocarbons (Table 1, Fig. 1). Because this decline is only ~3% of the total atmospheric burden of EECl, substantial ozone recovery will depend on further and continued decreases in the burden of contributing gases. When a recovery is first detected will depend also on the magnitude of 'natural' variability in column ozone and the future change of other variables that influence stratospheric ozone. These variables include stratospheric temperature, aerosol loading, and the abundance of trace gases that influence climate or the availability of reactive halogens in the stratosphere^{15–18} (for example, CH₄, N₂O).

An examination of observed rates of change for individual halocarbons in recent years demonstrates how the decline in EECl has resulted mainly from the rapid decrease in the atmospheric abundance of CH₃CCl₃ (Figs 1, 2a). But decreases in CH₃CCl₃ will not be sustained in coming years. The exponential rate of change was approximately –16% yr^{–1} in 1997, or near the limit expected for a global lifetime of 4.8 years (ref. 11). Accordingly, the absolute rate of change for CH₃CCl₃ is now near its maximum, and will become smaller by almost a factor of 3 over each 5-year period in the future even if emissions are completely eliminated after 1997 (Fig. 2). All other halocarbons that contribute significantly to the stratospheric burden of ozone-depleting halogens have longer atmospheric lifetimes and respond more slowly to changes in emissions. Unless emissions of these other halocarbons become smaller in the coming years, the rate of change for EECl will become substantially less negative in the near future (Fig. 2).

The importance of CH₃CCl₃ to the current decline in EECl is further underscored by noting that, without its influence, the sum of chlorine, or bromine, or EECl, would still have been increasing in 1997 (Fig. 2a). Atmospheric concentrations of hydrochlorofluorocarbons (HCFCs), halons and CCl₂F₂ (CFC-12) continue to increase and to augment the amount of Cl and Br in the atmosphere each year. In the absence of CH₃CCl₃, these increases would offset the small decreases currently observed for CCl₃F (CFC-11), CCl₂FCClF₂ (CFC-113) and CCl₄. Because the influence of CH₃CCl₃ has peaked and is now becoming smaller, EECl would stop decreasing by 2010 at only ~6% below its peak if emissions of ozone-depleting substances do not drop further over the next 10 years (Fig. 2b).

Table 1 Atmospheric halocarbon characteristics

Chemical	Industrial or common name	Lifetime* (yr)	Effective stratospheric release fraction†	f‡
CCl ₂ F ₂	CFC-12	100	0.48	1.06
CCl ₃ F	CFC-11	45	0.80	1.08
CCl ₂ FCClF ₂	CFC-113	85	0.60	1.06
CH ₃ CCl ₃	Methyl chloroform	4.8	0.86	1.09
CCl ₄	Carbon tetrachloride	35	0.85	1.09
CBBrF ₃	H-1301	65	0.64	1.08
CBBrClF ₂	H-1211	16	0.88	1.10
CHCl ₂ F	HCFC-22	12	0.28	1.10
CH ₃ CClF ₂	HCFC-142b	19	0.29	1.10
CH ₃ CCl ₂ F	HCFC-141b	9	0.58	1.15

* Steady-state lifetimes represent updates²³ to those in ref. 4. Although much uncertainty remains regarding the true atmospheric lifetime of H-1211, we use 16 yr here based on work²⁴ that accounts for photochemical loss in both the troposphere and stratosphere. Shorter lifetimes are more consistent with a budget analysis¹², but are less consistent with the observed global distribution.

† Effective stratospheric release fraction is an estimate of the fraction of a halocarbon destroyed in a single pass through the stratosphere. These estimates are from ref. 14 after multiplying by 0.8 to account for destruction of CFC-11 in a single pass through the stratosphere. EECl is calculated for each gas as follows: $EECl_x = X_x R_x (n_{Clx} + \alpha n_{Brx})$, where X_x is the mixing ratio of compound "x", R_x is the effective stratospheric release fraction, α is the equivalency factor applied to brominated gases (50 used here), and n_{Clx} and n_{Brx} represent the number of Cl and Br atoms per molecule in compound "x".

‡ Estimated from measured vertical gradients of halocarbons²³ or stratospheric lifetimes and allows for estimates of global burden from measurements made solely in the troposphere: $X_{global} = (X_{surface}/f)$ (ref. 25).

Continued increases are expected for HCFCs in the near future because these chemicals are used as interim replacements for ozone-depleting substances³. Their production is prohibited in developed countries after 2030. Understanding the continuing increases observed for halons and CFC-12, however, requires further consideration of the variables influencing the abundance of these trace gases in the atmosphere.

To investigate more clearly how human behaviour has affected the

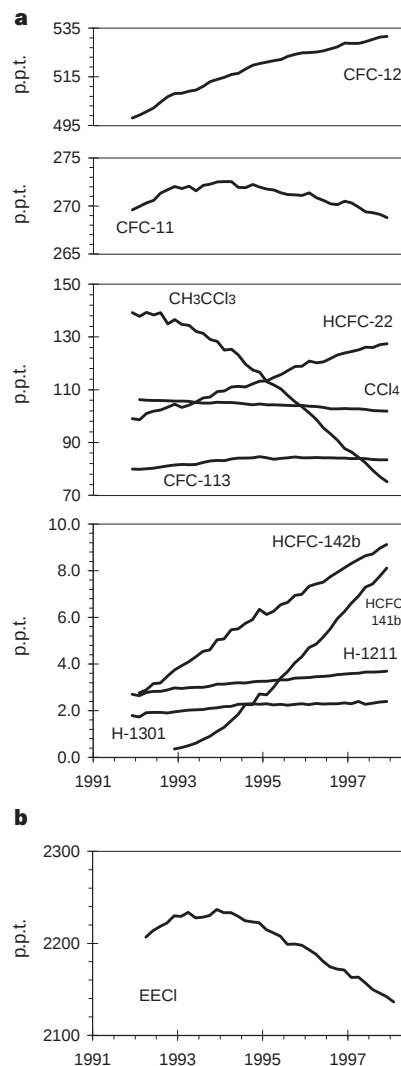


Figure 1 Measured mixing ratios. **a**, Halocarbons; **b**, the resulting effective equivalent chlorine, EECl. The global tropospheric mixing ratios are a weighted average of measurements at 5–7 remote sampling locations: Alert, North West Territories, Canada; Barrow, Alaska, USA; Niwot Ridge, Colorado, USA; Mauna Loa, Hawaii, USA; American Samoa; Cape Grim, Tasmania, Australia; and the South Pole. Gas chromatographs with electron capture detectors provide multiple measurements per day of CCl₃F (CFC-11), CCl₂F₂ (CFC-12), CH₃CCl₃, and CCl₄ at Barrow, Niwot Ridge, Mauna Loa, Samoa and the South Pole¹⁵. Measurements of CCl₂FCClF₂ (CFC-113), CBrClF₂ (halon-1211), CBrF₃ (halon-1301), CHClF₂ (HCFC-22), CH₃CCl₂F (HCFC-141b) and CH₃CClF₂ (HCFC-142b) are made from flask air samples collected at all 7 remote stations and analysed in Boulder, Colorado, with gas chromatographs having electron capture detectors or mass spectrometric detectors^{12,25–27}. Hemispheric mixing ratios are estimated after weighting results by sampling latitude, and these results are then averaged to estimate tropospheric, global means. We prepare and maintain calibration standards of all measured gases to minimize the influence of inaccuracy and calibration drift over time. Persistent chemicals not included in this figure (such as methyl halides and lesser halocarbons) contribute an additional 900 p.p.t. to EECl in today's atmosphere.

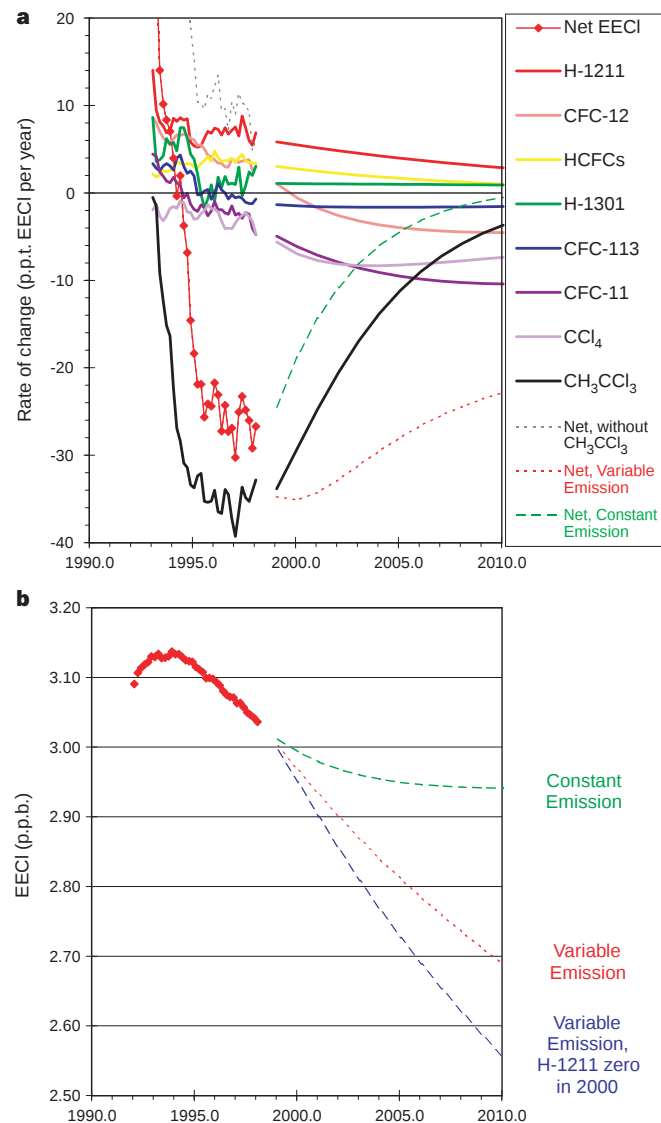


Figure 2 Trends in the rate of change and burden of EECl. **a**, The net rate of change in EECl and the contribution of individual halocarbons to this rate. Rates before 1998 are calculated as the difference in measured bimonthly mixing ratios during the preceding 12 months (for example, January 1997–January 1998 is plotted at January 1998) and conversion factors associated with calculating EECl (Table 1). With the exception of halons, HCFCs and CH₃CCl₃, future rates are calculated assuming continued emission declines and general compliance with the Montreal Protocol²¹. Emissions for halons and HCFCs are kept constant at 1997 rates to the end of 2010, and emissions of CH₃CCl₃ are assumed to be zero after 1997. The net result of these changes is labelled the “variable emission” scenario. Also shown is the past net rate of change for all chemicals excluding CH₃CCl₃ (EECl without CH₃CCl₃) and the potential future net rate of change if emissions of all halocarbons were to continue at 1997 rates (“constant emission” scenario, rates for individual halocarbons not shown). **b**, Measured and possible future EECl burden (parts per 10⁹, p.p.b.). Mixing ratios are plotted versus air sampling dates; these dates are relevant for the stratosphere only after considering time lags associated with mixing tropospheric air into the stratosphere (~3–5 yr). Points before 1998 are calculated from observations, lines after 1998 are calculated for emission scenarios described in **a** and the additional scenario, “variable emission, H-1211 zero in 2000”, which is the “variable emission” scenario except H-1211 emissions are eliminated in the year 2000. 900 p.p.t. has been added to EECl here to account for contributions of CH₃Cl, CH₃Br and other minor halogenated gases, for which mixing ratios are assumed to be constant over time. Similar results and trends were noted (but are not shown) for “equivalent chlorine”, which is more relevant for estimating an upper limit to the stratospheric halogen burden in polar regions in the springtime^{14,15} (“equivalent chlorine” is calculated as EECl with the exception that the effective stratospheric release fraction is 1.0 for all compounds).

atmospheric concentration of halocarbons in recent times we derive emissions of halocarbons with a simple model, using NOAA/CMDL atmospheric measurements and estimates of global removal rates (Fig. 3, Table 1). Results from this calculation indicate that emissions of most ozone-depleting substances were reduced substantially by 1997 compared to the late 1980s, a period when peak emissions occurred for most of these chemicals^{6–11,19,20} (Fig. 3). However, we note varying degrees of decline in emissions for the halocarbons. These differences can be explained in part by historic use patterns for these chemicals. For example, whereas solvents, compounds from fire extinguishers, cleaning agents and aerosol propellants are released to the atmosphere immediately on use, halocarbons employed as air-conditioning, refrigeration, and foam-blowing agents (the predominant uses for CFC-11 and CFC-12 in the past) are released more slowly over time.

As derived from NOAA/CMDL atmospheric measurements, emissions have decreased most dramatically for chemicals used previously as solvents and cleaning agents, that is, CFC-113 and CH₃CCl₃ (Fig. 3). The absence of significant quantities of these materials either as reserves^{8,9} or within in-use applications and the use of alternative chemicals and processes has allowed their emissions to decrease more rapidly than those of other gases. We also note that, although emissions of both these compounds have declined similarly, atmospheric mixing ratios of CH₃CCl₃ have diminished much more rapidly owing to its shorter lifetime (4.8 yr versus 85 yr; Table 1).

Emissions inferred for CCl₄ and the most abundant chlorofluorocarbons, CFC-11 and CFC-12, have decreased more slowly since 1992 (Fig. 3). Although similar relative declines are estimated for emissions of these three gases, the concentration of CFC-12 is still increasing because of its long atmospheric lifetime compared to CFC-11 and CCl₄ (Figs 1 and 3, Table 1). Slower declines in emissions of chlorofluorocarbons (CFCs) result in part because these chemicals were used predominantly in applications from which emission occurs over a number of years. From a simple accounting of industrial sales and emission estimates^{6,7}, it can be shown that significant amounts of these CFCs had been produced but not yet emitted to the atmosphere in 1995. We estimate that this reservoir or ‘bank’ of CFCs was large enough to sustain emissions at the 1995 rate for an additional 3–5 yr for CFC-11 and an additional 2 yr for CFC-12. Accordingly, even with a large decrease in production, emissions of CFCs will continue for a period from foams and cooling devices currently in use.

Any new production and use of these CFCs in recent years would also contribute to continued emissions. It has been suggested that CCl₄ emissions are closely tied to CFC production¹⁰. The relatively slow decline inferred for emissions of CCl₄ may suggest that CFCs are still being produced under the terms of the Montreal Protocol¹⁰ (Fig. 3). Despite the slower rate of decline for emissions of these CFCs compared with those of chemical solvents, the emissions inferred for 1992–1997 are at or below those expected based on overall adherence to the Montreal Protocol^{2,13,21}. Emissions deduced for CCl₄, however, are significantly above these expectations²¹.

Inferred emissions of hydrochlorofluorocarbons (HCFCs) increased slightly during 1992–97 as these compounds gained acceptance as interim replacements for CFCs and chlorinated solvents (Fig. 3). Molar emissions of HCFCs are now larger than the aggregate emission of CFC-11, CFC-12 and CFC-113. Emissions inferred for HCFCs from NOAA/CMDL measurements are approximately half of those estimated for compliance with the Montreal Protocol²¹, suggesting that production of HCFCs has been well below that allowed under the 1996 ‘production freeze’ for developed nations.

Halons are used almost exclusively as fire extinguishing agents for safeguarding and flooding of entire areas (CBrF₃ or H-1301), and for applying to spot-fires from portable devices (H-1211). Although production of these gases was to have been eliminated in developed

countries by January 1994, two years earlier than for CFCs, emissions inferred from measurements continued in the 1990s at significant rates (Fig. 3). Continued release of both halons results in part because substantial reservoirs of H-1211 and H-1301 exist in fire-extinguishing equipment today^{12,29}. Based on reported production of halon^{19,20,29}, this 'bank' is similar in size for both halons and large enough to sustain 1996 emissions of H-1211 for 8–12 years

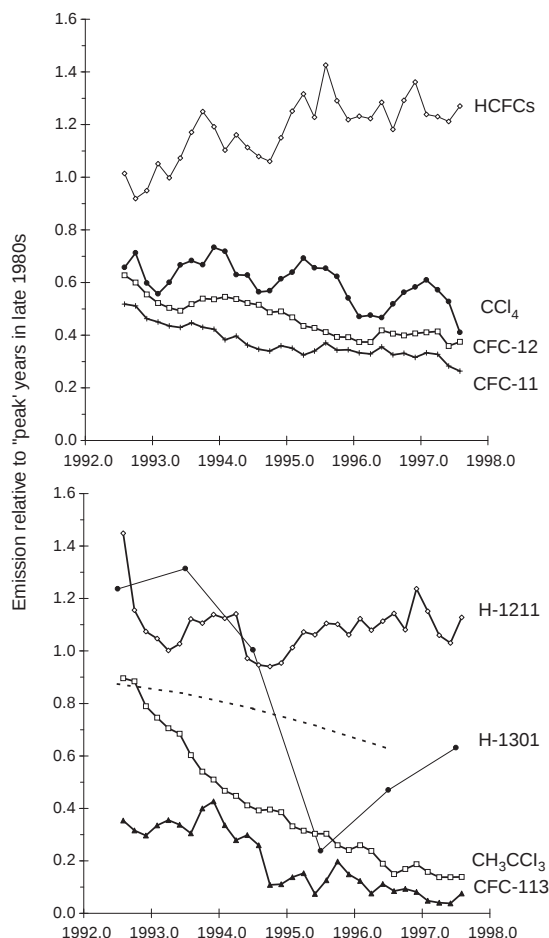


Figure 3 Halocarbon emissions derived from atmospheric measurements. These emissions have been normalized to "peak" emissions that occurred in the late 1980s. The results are displayed in two panels for clarity. Emissions are estimated from the observed change in bimonthly mixing ratios over the 12-month period centred around each datum, and the equation, $E = (M_a/f)(\Delta + x/\tau)$, where E is annual emission, M_a is the mass of the dry atmosphere²⁸, f is a unitless number that is used for converting tropospheric measurements to global atmospheric averages (Table 1), Δ is the observed, 12-month mixing-ratio difference, x is the observed, global tropospheric mean mixing ratio over the 12-month period, and τ is the mean, steady-state lifetime (Table 1), which is assumed to be constant over time. "Peak" emissions are calculated as the mean of the three peak years between 1986 and 1990 as inferred from measurements of H-1211 (ref. 12) and CCl_4 (ref. 10) and as estimated by industry for other gases^{6,7}. Also shown is the ratio inferred from industry emission estimates for H-1211 (refs 12, 19; dashed line). The peak annual emissions used are: CFC-11, 342 Gg; CFC-12, 450 Gg; CFC-113, 256 Gg; CH_3CCl_3 , 691 Gg; CCl_4 , 90 Gg; HCFC-22 + HCFC-141b + HCFC-142b, 207 Gg; H-1211, 9.0 Gg; H-1301, 3.8 Gg. Because of higher variability in measurements of H-1301, only annual means are plotted. Relative emission declines estimated here are reasonably consistent with emission estimates made elsewhere with different models and independent atmospheric measurements of CFC-113 (to the end of 1993)⁹, CFC-11 and CFC-12 (to the end of 1994)², and CCl_4 (to the end of 1995)¹⁰. Based on our measurements, the emission ratio inferred for H-1211 in 1996–1997 ($E_{1996-1997}/E_{\text{peak}}$) is 1.11 (+0.15, –0.13). The uncertainty in this ratio is calculated from the sum of squares of variability in measurements, a lifetime range of 11–24 yr, potential calibration drift of $\pm 5\%$, and potential lifetime transients of $\pm 5\%$.

into the future and 1996 emissions of H-1301 for another 40 years (refs 12, 29). Emissions from the halon 'bank', however, are different from emission of CFCs. Whereas CFCs are released passively from in-use applications, halons are emitted by active participation of personnel involved in fire-fighting, training, or servicing of fire equipment²⁰. Adherence to recent protocols for reducing unnecessary emissions of halons during equipment servicing and personnel training ought to have reduced atmospheric accumulation rates for these chemicals, because emissions from use on fires is thought to have contributed only $\sim 10\%$ of past emissions relative to these other uses²⁰. This probably explains the slower accumulation rate and reduced emissions of H-1301 in recent years despite the presence of a substantial quantity of this halon in reserve^{12,29} (Figs 2 and 3).

Similar emission reductions have not been realized for H-1211 in recent years. Annual emissions inferred for this halon during 1992–97 are similar to those estimated for the late 1980s (Fig. 3), a pattern not dissimilar from that observed for the HCFCs in recent years. These observations indicate that emissions of H-1211 have continued despite efforts to halt production, to limit use to only those deemed essential, and to recover and recycle halon²². Furthermore, emissions of H-1211 continue despite the availability of alternatives for most of its uses²⁰.

Because emissions have not dropped and because it contains Br, H-1211 has retarded the decline of EECl more than any other halocarbon in recent years (Fig. 2a). Continued emission of H-1211 at 1997 rates would result in a total burden of EECl that is higher by as much as 3–5% during the next 5–10 years than if emissions were eliminated in the year 2000 (Fig. 2b).

In contrast to the conclusion drawn from atmospheric measurements, industry estimates^{12,29} suggest that emissions of H-1211 should have declined in the 1990s (Fig. 3). Sustained emissions of halons may result from use of reserves at a more rapid rate than estimated by industry, or from production and emission in excess of that included in current models. Production in developing nations is allowed under the Montreal Protocol until 2010 (ref. 3). Production of H-1211 has been reported so far by India, the Korean Republic and China in recent years²⁰. Although reported production is included in emission models^{12,29}, models that use different algorithms²⁰ to compute emissions could predict a relatively constant emission rate in the 1990s, as is suggested by the atmospheric observations.

As to the future, once halocarbon emissions become insignificant, the decline in EECl will be determined solely by the residence times of these chemicals in the atmosphere. Although the largest emissions reductions have been noted for CH_3CCl_3 and CFC-113, emissions of other gases continue at substantial rates. For some chemicals, in particular H-1211, increased efforts to reduce emissions further could make a substantial difference in how EECl drops in the future. If halocarbon emissions were to remain constant into the future at 1997 rates, the burden of reactive halogens (EECl) would stop decreasing by about 2010 as the influence of CH_3CCl_3 diminishes (Fig. 2b). Although this is not expected if countries adhere to the Montreal Protocol, it emphasizes that stratospheric ozone recovery will be achieved only if halocarbon emissions continue to decline. □

Received 3 September 1998; accepted 5 February 1999.

- Montzka, S. A. *et al.* Decline in the tropospheric abundance of halogen from halocarbons: Implications for stratospheric ozone depletion. *Science* **272**, 1318–1322 (1996).
- Cunnold, D. M. *et al.* GAGE/AGAGE measurements indicating reductions in global emissions of CCl_3F and CCl_2F_2 in 1992–1994. *J. Geophys. Res.* **102**, 1259–1269 (1997).
- Report of the Ninth Meeting of the Parties to the Montreal Protocol on Substances that Deplete the Ozone Layer (Montreal) (United Nations Environmental Programme, New York, 1997).
- Scientific Assessment of Ozone Depletion: 1994 (Rep. No. 37, World Meteorological Organization, Geneva, 1995).
- Elkins, J. W. *et al.* Decrease in the growth rates of atmospheric chlorofluorocarbons 11 and 12. *Nature* **364**, 780–783 (1993).
- Production, Sales and Atmospheric Release of Fluorocarbons Through 1995 (AFEAS, Washington DC, 1997).
- Fisher, D. A. *et al.* in *Report on Concentrations, Lifetimes, and Trends of CFCs, Halons, and Related Species* Ch. 2 (NASA ref. publ. 1339, Washington DC, 1994).
- Midgley, P. M. & McCulloch, A. The production and global distribution of emissions to the atmosphere of 1,1,1-trichloroethane (methyl chloroform). *Atmos. Environ.* **29**, 1601–1608 (1995).

9. Fraser, P. *et al.* Lifetime and emission estimates of 1,1,2-trichlorotrifluoroethane (CFC-113) from daily global background observations June 1982–June 1994. *J. Geophys. Res.* **101**, 12585–12599 (1996).
10. Simmonds, P. G. *et al.* Global trends and emission estimates of CCl₄ from in situ background observations from July 1978 to June 1996. *J. Geophys. Res.* **103**, 16017–16028 (1998).
11. Prinn, R. G. *et al.* Atmospheric trends and lifetime of CH₃CCl₃ and global OH concentrations. *Science* **269**, 187–192 (1995).
12. Butler, J. H. *et al.* Growth and distribution of halons in the atmosphere. *J. Geophys. Res.* **103**, 1503–1511 (1998).
13. Prather, M. J. & Watson, R. T. Stratospheric ozone depletion and future levels of atmospheric chlorine and bromine. *Nature* **344**, 729–734 (1990).
14. Daniel, J. S., Solomon, S. & Albritton, D. L. On the evaluation of halocarbon radiative forcing and global warming potentials. *J. Geophys. Res.* **100**, 1271–1285 (1995).
15. Jackman, C. H. *et al.* Past, present, and future modeled ozone trends with comparisons to observed trends. *J. Geophys. Res.* **101**, 28753–28767 (1996).
16. Portmann, R. W. *et al.* Role of aerosol variations in anthropogenic ozone depletion in the polar regions. *J. Geophys. Res.* **101**, 22991–23006 (1996).
17. Solomon, S. *et al.* The role of aerosol variations in anthropogenic ozone depletion at northern midlatitudes. *J. Geophys. Res.* **101**, 6713–6727 (1996).
18. Shindell, D. T., Rind, D. & Loneragan, P. Increased polar stratospheric ozone loss and delayed eventual recovery owing to increasing greenhouse-gas concentrations. *Nature* **392**, 589–592 (1998).
19. McCulloch, A. Global production and emissions of bromochlorodifluoromethane and bromotrifluoromethane (halons 1211 and 1301). *Atmos. Environ.* **A 26**, 1325–1329 (1992).
20. UNEP Technology and Economic Assessment Panel April 1998 Report (United Nations Environment Programme, Nairobi, Kenya, 1998); also available at (http://www.teap.org/html/teap_reports.html).
21. Prather, M. J. *et al.* in *Climate Change 1995, the Science of Climate Change* (eds Houghton, J. T. *et al.*) Ch. 33 (Cambridge Univ. Press, 1996).
22. Report of the Halon Fire Extinguishing Agents Technical Options Committee (United Nations Environment Programme, Nairobi, Kenya, 1994).
23. Volk, C. M. *et al.* Evaluation of source gas lifetimes from stratospheric observations. *J. Geophys. Res.* **102**, 25543–25564 (1997).
24. Burkholder, J. B. *et al.* Atmospheric fate of CF₃Br, CF₂Br₂, CF₂ClBr, and CF₂BrCF₂Br. *J. Geophys. Res.* **96**, 5025–5043 (1991).
25. Butler, J. H. *et al.* A decrease in the growth rates of atmospheric halon concentrations. *Nature* **359**, 403–405 (1992).
26. Montzka, S. A. *et al.* Global tropospheric distribution and calibration scale of HCFC-22. *Geophys. Res. Lett.* **20**, 703–706 (1993).
27. Montzka, S. A. *et al.* Early trends in the global tropospheric abundance of hydrochlorofluorocarbon-141b and -142b. *Geophys. Res. Lett.* **21**, 2483–2486 (1994).
28. Trenberth, K. E. & Guillemot, C. J. The total mass of the atmosphere. *J. Geophys. Res.* **99**, 23079–23088 (1994).
29. Fraser, P. J. *et al.* Southern hemispheric halon trends (1978–1998) and global halon emissions. *J. Geophys. Res.* (in the press).

Acknowledgements. We thank all personnel involved in collecting flask samples at the NOAA/CMDL Observatories and at cooperative sampling sites. S.A.M. appreciates discussions and suggestions from S. Solomon and P. Midgley, and we appreciate the past technical assistance of R. C. Myers and T. H. Swanson. This work was supported in part by NOAA's Radiatively Important Trace Species research and by the Atmospheric Chemistry project of the NOAA Climate and Global Change Program.

Correspondence and requests for materials should be addressed to S.A.M. (e-mail: smontzka@cmdl.noaa.gov). Data used in this Letter are available at (<http://www.cmdl.noaa.gov/ftpdata.html>) or (<ftp://ftp.cmdl.noaa.gov/noah>).

Increased El Niño frequency in a climate model forced by future greenhouse warming

A. Timmermann, J. Oberhuber*, A. Bacher, M. Esch, M. Latif & E. Roeckner

Max-Planck-Institut für Meteorologie und * Deutsches Klimarechenzentrum, Bundesstrasse 55, D-20146 Hamburg, Germany

The El Niño/Southern Oscillation (ENSO) phenomenon is the strongest natural interannual climate fluctuation¹. ENSO originates in the tropical Pacific Ocean and has large effects on the ecology of the region, but it also influences the entire global climate system and affects the societies and economies of many countries². ENSO can be understood as an irregular low-frequency oscillation between a warm (El Niño) and a cold (La Niña) state. The strong El Niños of 1982/1983 and 1997/1998, along with the more frequent occurrences of El Niños during the past few decades, raise the question of whether human-induced 'greenhouse' warming affects, or will affect, ENSO³. Several global climate models have been applied to transient greenhouse-gas-induced warming simulations to address this question^{4–6}, but the results have been debated owing to the inability of the models to fully simulate ENSO (because of their coarse equatorial resolution)⁷. Here we present results from a global climate

model with sufficient resolution in the tropics to adequately represent the narrow equatorial upwelling and low-frequency waves. When the model is forced by a realistic future scenario of increasing greenhouse-gas concentrations, more frequent El Niño-like conditions and stronger cold events in the tropical Pacific Ocean result.

Our global climate model^{8,9} uses a meridional resolution of 0.5° in the tropics. The model, which is 'flux corrected', simulates an irregular ENSO cycle similar to the observed one, and the amplitude and dynamics of the simulated ENSO cycle are consistent with those derived from observations^{8,9}. The simulated ENSO period is too short, however, and amounts to about two years, whereas observations indicate a main period of about four years. Our model successfully predicted the onset and decline of the 1997/1998 El Niño several months in advance¹⁰.

Here, two experiments were performed. The first experiment is a 300-year-long control integration with fixed present-day atmospheric concentrations of greenhouse gases. The second experiment is a transient greenhouse warming simulation in which the model was forced by increasing levels of greenhouse gases as observed (1860–1990) and according to IPCC scenario IS92a¹¹ (1990–2100).

The changes in the mean state at the surface of the tropical Pacific Ocean, as derived from the transient greenhouse warming simulation, are reminiscent of the anomalous climate state observed during present-day El Niño conditions (Fig. 1). The sea-surface-temperature (SST) trend pattern is characterized by strongest warming in the equatorial east Pacific, accompanied by westerly near-surface wind anomalies in the equatorial region to the west of the maximum warming and strong equatorward flow off the Equator (not shown). The associated trend in rainfall is rather similar to that simulated during present-day El Niños (not shown).

There has been some discussion about the relative roles of different feedbacks involved in the time-mean response of the tropical Pacific climate system to greenhouse warming. On the one hand, it has been suggested that regional differences in the cloud-albedo feedback will lead to surface warming that is strongest in the equatorial east Pacific¹². The argument is that the equatorial west Pacific is so warm that even modest additional warming would lead to a cloud shielding effect, with high cirrus clouds, reducing incoming solar radiation at the surface and inhibiting further warming¹³. This 'thermostat' would be less efficient in the eastern equatorial Pacific, so it would warm more than the western Pacific. This would lead to a slackening of the winds along the Equator and

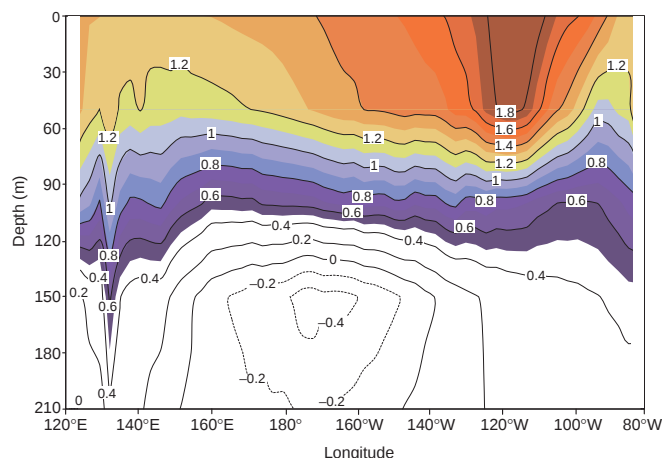


Figure 1 Simulated temperature trends in equatorial waters. The linear trends in the temperatures of the upper 210 m of the Pacific Ocean at the Equator (°C rise per 100 years) are derived from the full (240-year-long) transient greenhouse warming simulation. There is a warming trend near the surface and a cooling trend at deeper levels, leading to a stronger thermocline. The trends at the surface resemble the anomalous conditions observed during present-day El Niños.

result in overall conditions very similar to those observed during El Niños.

On the other hand, it has been argued that the strong equatorial upwelling in the eastern equatorial Pacific will weaken the warming in this region, so the strongest warming will occur in the western equatorial Pacific⁷. This would lead to stronger winds along the Equator, more equatorial upwelling and a net cooling in the eastern Pacific. This 'dynamical thermostat' will lead to overall conditions resembling those observed during La Niñas, eventually retarding global warming. The global climate models applied so far to greenhouse warming simulations could not adequately address this problem⁷ as they were poor at resolving the equatorial upwelling. Our climate model includes adequate representations of both types of feedback, and the balance of all physical processes leads to a simulated warming pattern that is strongest in the east, consistent with ref. 12.

As can be seen from the time evolution of the eastern equatorial SST in the transient greenhouse warming simulation, there is considerable interannual variability superimposed on the warming trend in the tropical Pacific (Fig. 2). Our model simulates an irregular ENSO cycle under enhanced greenhouse conditions, with a main period close to that derived from the control integration. However, the ENSO cycle evolves under different mean conditions to those currently found so we may expect the statistics of the ENSO cycle to change under enhanced greenhouse conditions. It can be seen from the SST time series (Fig. 2) that the level of the interannual variability increases strongly towards the end of the greenhouse warming simulation. This is also seen in the time series of the interannual SST standard deviations, which shows a strong increase towards the end of the transient greenhouse warming integration (Fig. 3). The observations also show that the interannual variability has intensified during the past several decades (Fig. 3), but this is well within the range of our control simulation.

The statistical significance of the increase in ENSO variability is an important issue to be addressed. In addition to our 300-year

control integration, we can use the first 100 years of the scenario integration in the estimation of the natural variability as the external forcing is small up to 1960. We also performed two scenario integrations that included anthropogenic sulphate aerosols¹⁴, which started in 1860 and were integrated up to the year 2050 (not shown). Although these two runs are too short to study changes in the statistics of the interannual variability, their first 100 years can be used in the estimation of the noise, so we have 600 years in total. We did not find any period within the whole 600 years that exceeds a running standard deviation of 1.15 °C. The enhanced ENSO variability seen towards the end of our greenhouse warming integration (Fig. 3) is therefore highly significant.

To investigate further the changes in the ENSO statistics, we computed the frequency distributions of monthly SST anomalies. The distribution obtained from the first half of the transient integration is narrower than that obtained from the second half, so the year-to-year variability becomes more extreme under enhanced greenhouse conditions. Furthermore, although the distributions of the SST anomalies in the control integration (not shown) and during the first half of the transient integration are almost symmetric (Fig. 2c), the distribution for the second half of the transient integration is skewed: strong cold extremes become more frequent, while the statistics of the strong warm extremes do not change (Fig. 2d).

Which changes in the mean state lead to changes in the statistics of the interannual variability? First, the changes in the mean state near the surface (Fig. 1) could favour a reduction in ENSO-type variability as the zonal asymmetries across the equatorial Pacific are reduced¹⁵. Second, interactions between the air and the sea may be more energetic in a warmer climate, increasing the interannual variability. Third, changes in the vertical density structure of the ocean may alter the level of the interannual variability. The equatorial thermocline becomes stronger in response to greenhouse warming (Fig. 1): temperatures near the surface rise, but those at deeper ocean levels fall. This cooling at subsurface levels can be

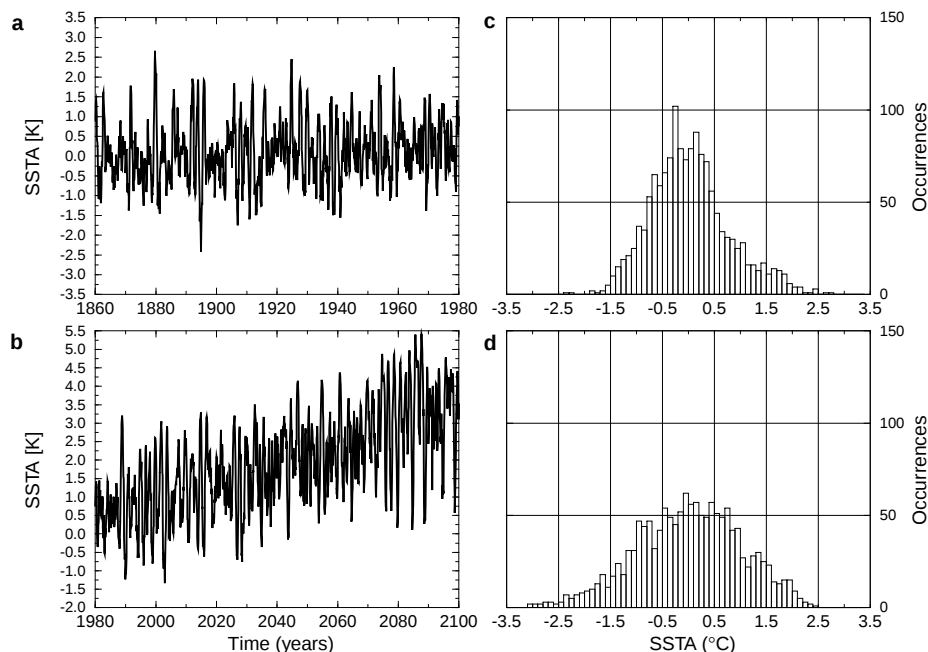


Figure 2 Simulated sea-surface-temperature anomalies. **a, b**, Time series of eastern equatorial SST anomalies (SSTA; relative to the control run) averaged over the Niño-3 region (150°W–90°W, 5°N–5°S) (°C) during the first half (**a**) and second half (**b**) of the transient greenhouse warming simulation. **c, d**, Frequency distribution of Niño-3 SST anomalies during the first half (**c**) and second half (**d**) of the transient greenhouse warming simulation. Anomalies were computed by subtracting the linear trends and means from the time series of eastern equatorial

Pacific SSTs and by removing the mean annual cycle of the control integration. Note the change in the frequency distributions from the first to the second half of the transient integration. The distribution is almost symmetric (with a small bias towards the warm side) during the first half (and during the control integration; not shown), but is biased towards the cold side during the second half of the transient integration: strong cold events become more frequent.

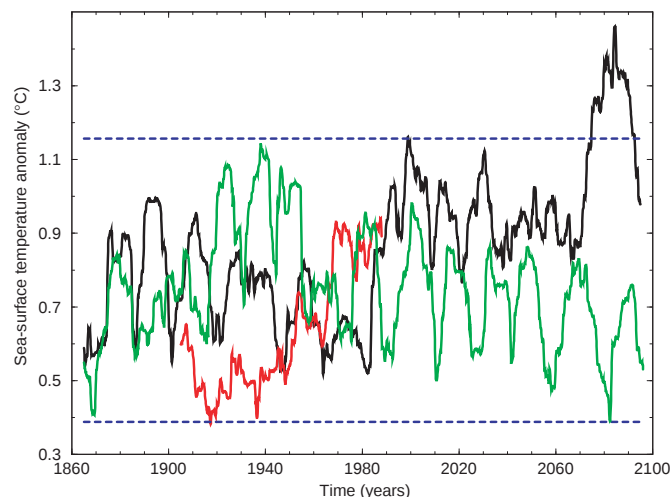


Figure 3 Interannual variability of observed and simulated SST anomalies. Standard deviations of Niño-3 SST anomalies ($^{\circ}\text{C}$) as function of time during the transient greenhouse warming simulation (black). Also shown are the time evolutions of the standard deviation of the observed (red) and control run (green) Niño-3 SST anomalies. A low-pass filter in the form of a sliding window 10 years wide was used to compute the standard deviations. Both the simulated and observed SST anomalies exhibit trends towards stronger interannual variability, with pronounced interdecadal variability superimposed. The minimum and maximum standard deviations derived from the control run with present-day concentrations of greenhouse gases are denoted by the dashed lines.

attributed to a greater inflow of cold waters in response to the intensification of the atmospheric Hadley Circulation, especially in the Southern Hemisphere. In order to gain further insight into the dynamics of the changes in the ENSO statistics, we computed atmospheric and oceanic sensitivities as functions of time (Fig. 4). The most important atmospheric forcing for equatorial oceans is the zonal wind stress component. We computed first the sensitivity of central equatorial zonal wind stress anomalies to the SST anomalies in the eastern equatorial Pacific averaged over the Niño-3 region (150°W – 90°W , 5°N – 5°S). No significant change was found. We then computed the sensitivity of Niño-3 SST anomalies to changes in the central equatorial zonal wind stress. This exhibited a significant increase towards the end of the transient greenhouse warming simulation, indicating that changes in the ocean dynamics arising from the strengthening thermocline are responsible for the enhanced interannual variability. This result is consistent with findings obtained from simpler coupled models^{16–18}.

The simple models also predict a skew in the frequency distribution of the thermocline depth anomalies in the eastern equatorial Pacific (a quantity closely related to eastern equatorial SST anomalies) for a sharpening thermocline, with strong cold events becoming more frequent¹⁸. This is consistent with our greenhouse warming simulation (Fig. 2). We therefore conclude that the most important change in the mean state of the tropical Pacific Ocean–atmosphere system affecting the ENSO statistics appears to be the strengthening of the equatorial thermocline.

The tropical Pacific climate system is thus predicted to undergo strong changes if emissions of greenhouse gases continue to increase. The climatic effects will be threefold. First, the mean climate in the tropical Pacific region will change towards a state corresponding to present-day El Niño conditions. It is therefore likely that events typical of El Niño will also become more frequent. Second, a stronger interannual variability will be superimposed on the changes in the mean state, so year-to-year variations may become more extreme under enhanced greenhouse conditions. Third, the interannual variability will be more strongly skewed, with strong cold events (relative to the warmer mean state)

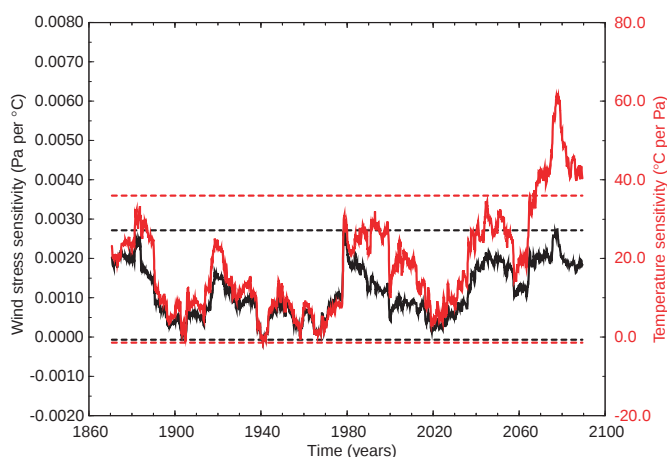


Figure 4 Atmospheric and oceanic model sensitivity. Sensitivity of central equatorial zonal wind stress anomalies (averaged over the region 100°W – 160°W , 1.4°N – 1.4°S) to changes in the Niño-3 SST anomalies as function of time during the transient greenhouse warming simulation (black). The atmospheric sensitivity is defined as the covariance of the zonal wind stress and Niño-3 SST anomalies divided by the variance of the Niño-3 SST anomalies ($\text{Pa per } ^{\circ}\text{C}$). Also shown is the sensitivity of eastern equatorial SST anomalies averaged over the Niño-3 region to the central equatorial zonal wind stress anomalies averaged over the region indicated above (red). The oceanic sensitivity is defined as the covariance of the zonal wind stress and Niño-3 SST anomalies divided by the variance of the zonal wind stress anomalies ($^{\circ}\text{C per Pa}$). A sliding window 10 years wide was used to compute the sensitivities. The minima and maxima derived from the control run are shown by the dashed black and red lines, respectively. Only the temperature sensitivity exhibits a statistically significant increase, indicating that the changes in the ocean dynamics in response to the strengthening thermocline are responsible for the increase in the level of the interannual variability.

becoming more frequent. Although the model was successful in simulating and predicting ENSO, the ENSO response to greenhouse warming may depend on processes that are not well understood, such as cloud feedback, so we cannot exclude the possibility that the results are sensitive to the model formulation. $\square$

Received 21 December 1998; accepted 22 February 1999.

- Philander, S. G. H. *El Niño, La Niña, and the Southern Oscillation* (Academic, San Diego, 1990).
- Glanzt, M. H., Katz, R. W. & Nicholls, N. *Teleconnections Linking Worldwide Climate Anomalies* (Cambridge Univ. Press, 1991).
- Trenberth, K. & Hoar, T. The 1990–1995 El Niño–Southern Oscillation event: Longest on record. *Geophys. Res. Lett.* **23**, 57–60 (1996).
- Knutson, T. R., Manabe, S. & Gu, D. Simulated ENSO in a global coupled ocean–atmosphere model: multidecadal amplitude modulation and CO_2 sensitivity. *J. Clim.* **10**, 138–161 (1997).
- Tett, S. Simulation of El Niño–Southern Oscillation-like variability in a global AOGCM and its response to CO_2 increase. *J. Clim.* **8**, 1473–1502 (1995).
- Meehl, G. A., Branstator, G. W. & Washington, W. M. Tropical Pacific interannual variability and CO_2 climate change. *J. Clim.* **6**, 42–63 (1993).
- Cane, M. A. *et al.* Twentieth-century sea surface temperature trends. *Science* **275**, 957–960 (1997).
- Roegner, E., Oberhuber, J. M., Bacher, A., Christoph, M. & Kirchner, I. ENSO variability and atmospheric response in a global atmosphere–ocean GCM. *Clim. Dynam.* **12**, 737–754 (1996).
- Bacher, A., Oberhuber, J. M. & Roegner, E. ENSO dynamics and seasonal cycle in the tropical Pacific as simulated by the ECHAM4/OPYC3 coupled general circulation model. *Clim. Dynam.* **14**, 431–450 (1997).
- Oberhuber, J. M., Roegner, E., Christoph, M., Esch, M. & Latif, M. Predicting the '97 El Niño event with a global climate model. *Geophys. Res. Lett.* **25**, 2273–2276 (1998).
- Houghton, J. T., Callander, B. A. & Varney, S. K. V. (eds) *IPCC Climate Change 1992. The Supplementary Report to the IPCC Scientific Assessment* (Cambridge Univ. Press, 1992).
- Meehl, G. A. & Washington, W. M. El Niño-like climate change in a model with increased atmospheric CO_2 concentrations. *Nature* **382**, 56–60 (1996).
- Ramanathan, V. & Collins, W. Thermodynamic regulation of ocean warming by cirrus clouds deduced from observations of the 1987 El Niño. *Nature* **351**, 27–32 (1991).
- Roegner, E., Bengtsson, L., Feichter, J., Lelieveld, J. & Rodhe, H. Transient climate change simulations with a coupled atmosphere–ocean GCM including the sulfur cycle. *J. Clim.* (in the press).
- Battisti, D. S. & Hirst, A. C. Interannual variability in the tropical atmosphere/ocean system: Influence of the basic state, ocean geometry and non-linearity. *J. Atmos. Sci.* **46**, 1687–1712 (1989).
- Münich, M., Cane, M. A. & Zebiak, S. E. A study of self-excited oscillations of the tropical ocean–atmosphere system, Part II: Nonlinear cases. *J. Atmos. Sci.* **48**, 1238–1248 (1991).
- Tziperman, E., Stone, L., Cane, M. A. & Jarosh, H. El Niño chaos: overlapping of resonances between the seasonal cycle and the Pacific Ocean–Atmosphere Oscillator. *Science* **264**, 72–74 (1994).
- Stone, L., Saparin, P. I., Huppert, A. & Price, C. El Niño chaos: The potential role of noise and stochastic resonance on the ENSO cycle. *Geophys. Res. Lett.* **25**, 175–178 (1998).

Acknowledgements. We thank T. P. Barnett for discussion and M. Münnich for help with data processing. This work was sponsored by the German government under its programme 'Klimavariabilität und Signalanalyse' and by the European Union through its 'SINTEX' programme. The climate model integrations were performed at the Deutsches Klimarechenzentrum.

Correspondence and requests for materials should be addressed to M.L. (e-mail: latif@dkrz.de).

Continental-shelf sediment as a primary source of iron for coastal phytoplankton

Kenneth S. Johnson^{*†}, Francisco P. Chavez[†] & Gernot E. Friederich[†]

^{*} Moss Landing Marine Laboratories, PO Box 450, Moss Landing, California 95039, USA

[†] Monterey Bay Aquarium Research Institute, PO Box 628, Moss Landing, California 95039, USA

The availability of iron, an essential nutrient, controls rates of phytoplankton primary productivity in the open-ocean, upwelling ecosystems of the equatorial Pacific^{1,2}. Upwelling injects large amounts of macronutrients into the euphotic zone of eastern boundary currents, such as the California Current System (CCS), where iron can become the limiting factor on productivity^{3,4}. Iron addition to samples from some areas of the CCS has been shown to increase rates of biomass production^{5,6}, but the processes that control iron availability in these systems remain poorly understood. Here we report measurements of dissolvable iron (that is, dissolved plus leachable iron at pH 3) in transects across the CCS in March of 1997 and 1998. We found high concentrations of iron in 1997 during strong upwelling conditions. During the 1998 El Niño, the concentration of dissolvable iron in surface waters was low, even though that year was marked by high river flow and low offshore salinity. These results indicate that the primary source of iron in the CCS is resuspension of particles in the benthic boundary layer, followed by upwelling of this iron-rich water, rather than direct riverine input. This source of iron must be an essential but variable component of the high productivity found in upwelling ecosystems.

Upwelling along eastern boundary currents sustains some 11% of the global ocean primary production⁷ and 50% of the global fish production⁸. The waters that come to the surface in these systems are enriched in macronutrients. However, there is no corresponding subsurface enrichment of iron in the offshore areas that are the source of the upwelled waters³. Consequently, there must be additional sources of iron at the continental margin to sustain elevated rates of primary production. Large concentration gradients of dissolved and particulate iron across both eastern⁹ and western¹⁰ ocean boundaries demonstrate that ocean margins are important sources of iron. It has been hypothesized that elevated iron concentrations in coastal regions may arise from riverine input, resuspended sediment or atmospheric deposition⁹, but so far we do not know the relative importance of these mechanisms or how they vary over time.

We have made measurements of dissolvable iron, nitrate and chlorophyll in March 1997 (cruise S197) and March 1998 (cruise S298) at a depth of 2 m along transects from Monterey Bay in a southwesterly direction across the CCS (Fig. 1). Conditions in March of 1997 were representative of typical spring upwelling, with strong winds from the northwest (Fig. 2c). A sharp drop in surface temperature, driven by strong winds that favoured upwelling, was recorded just before the S197 cruise at the MBARI M1 mooring in Monterey Bay¹¹ (Fig. 2a). Temperatures remained low throughout the cruise and dissolved nitrate concentrations near the

surface were as high as 25 μM . Surface salinities measured along the transect were >32.8 , which is typical of the CCS in this region¹². The shipboard measurements of dissolvable iron at stations within 100 km of the coast averaged 6 nM, but values as high as 20 nM were observed near the core of the upwelling system at Año Nuevo, which feeds Monterey Bay¹³. In the open ocean waters more than 200 km offshore (west of 124.2° W), the dissolvable iron concentration was 0.2 ± 0.1 (mean $\pm$ standard deviation) nM. High chlorophyll concentrations developed near the coast under these conditions (Fig. 1c).

Wind conditions that favoured upwelling were rare during the El Niño conditions of late winter and early spring of 1998. High-wind events were primarily from the south (Fig. 2c), which would have suppressed offshore transport, and the surface ocean was nearly 3 °C

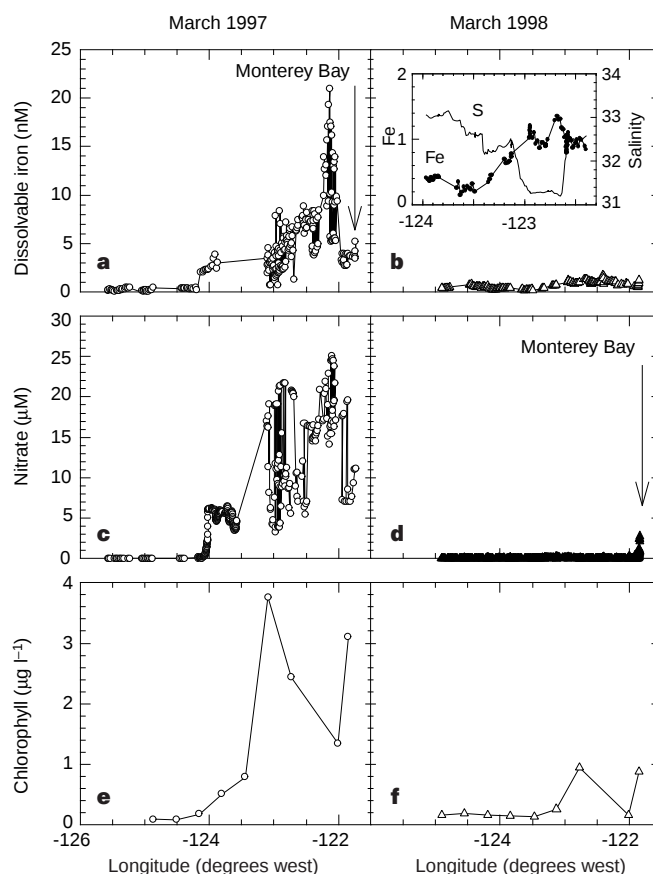


Figure 1 Dissolvable iron, nitrate and chlorophyll on transects across the CCS. Measurements were made from Moss Landing in Monterey Bay (36.80° N, 121.79° W) to a point 390 km offshore (35.07° N, 125.57° W) on cruise S197 (3–9 March 1997) and 317 km offshore (35.46° N, 124.90° W) on cruise S298 (March 18–23 1998). The inset to **b** shows iron concentration on an expanded scale and the salinity on S298, which is anomalously low over the inner 150 km. Dissolvable iron was measured by flow-injection analysis with chemiluminescence detection at 6-min intervals²⁴. Sea water was sampled with an all-tylon pump system towed at 1–2 m depth. Nitrate in this sample stream was determined colorimetrically aboard ship using an automated continuous-flow analysis system²⁵. Sea water from the pump was passed through a 10- μm polypropylene filter and acidified to pH 3 for ~0.5 min before iron analysis. Laboratory experiments demonstrate no difference between the amount of iron desorbed from particles at pH 3 in 1 min and in >6 h²⁶, indicating that we detected dissolved plus all absorbed iron. The dissolvable iron/nitrate relationship in the data was similar to that reported previously for Monterey Bay³. Chlorophyll was determined by extracting the pigment and measuring fluorescence in a shore-based laboratory. The nitrate data near the upwelling front at 124° W on S197 were obtained on the inshore leg of the cruise to replace data missing on the offshore transect.

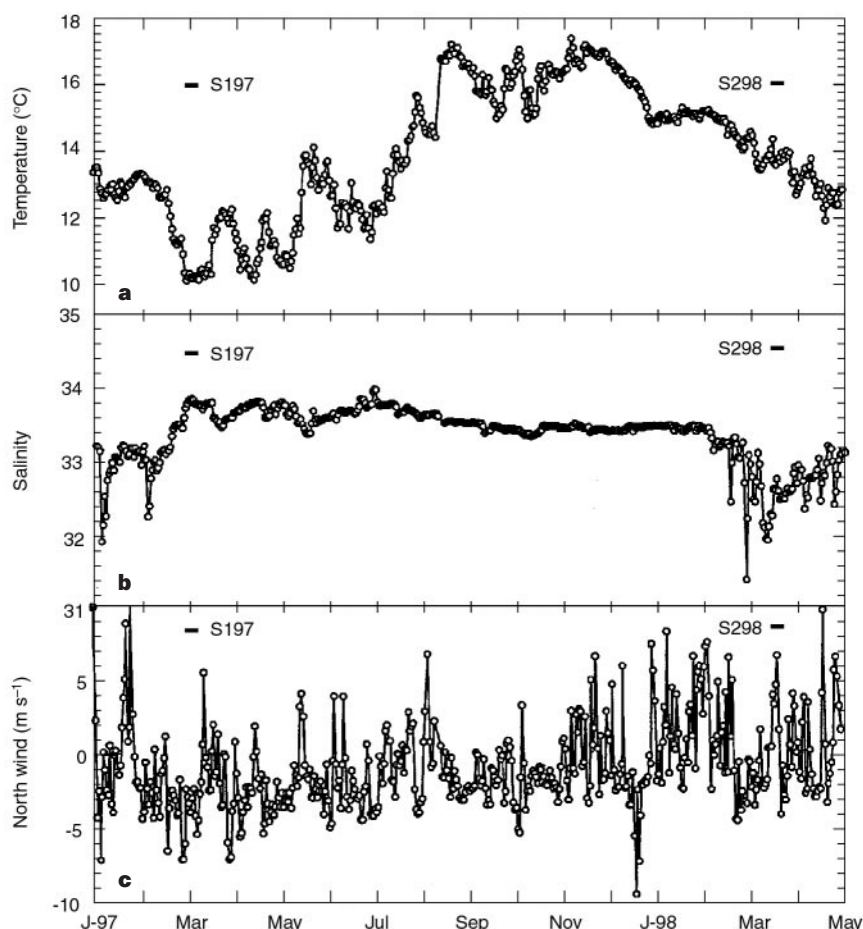


Figure 2 Daily average values of temperature, salinity and north windspeed. Parameters were measured from 1 January 1997 to 30 May 1998 on the M1 mooring in Monterey Bay (36.73° N, 122.01° W). The time spans of the S197 and

S298 cruises are shown. Data are available from the MBARI database located on the Internet at www.mbari.org.

warmer than in 1997 (Fig. 2a). Rainfall during the winter of 1997/98 was double the annual average at meteorological stations throughout the central California region. Salinity across the inner 150 km of this transect was less than 32.8 to depths of 10–15 m (Figs 1b, 2b), which is anomalously low¹². Drifting buoys drogued with 10-m holey socks of the WOCE design and equipped with ARGOS satellite transmitters were deployed at three stations along the cruise track. The inner two drifters deployed near 122.5° and 123.8° W moved continuously south for the next month at an average speed of 10 cm s⁻¹. The mean speed increased to 20 cm s⁻¹ when the wind was from the north. The drifter tracks show that the low salinities must have been produced by outflow from the Sacramento/San Joaquin River system that has been carried south in the CCS. These two rivers, which discharge 100 km to the north of the transect through San Francisco Bay, combine to form one of the 60 largest systems in the world. The dissolvable iron concentrations in the offshore waters were slightly higher (0.4 ± 0.1 nM) than the values recorded in 1997, whereas values in the river-influenced inner 100 km of the transect had an average value of 0.9 ± 0.2 nM (Fig. 1b). Little chlorophyll was found in this system except in plumes of low-salinity water at 122.8° W and 121.8° W (Fig. 1f).

Our measurements confirm previous observations that iron concentrations in coastal systems can be elevated, relative to open-ocean concentrations^{9,10}. However, the high-resolution ship-board measurements also provide significant new information on the mechanism by which iron is introduced in coastal systems; this information cannot be obtained by classical methods for iron analysis⁹. Periods of low salinity from freshwater runoff regularly occur each winter along the central California coast¹⁴ (Fig. 2b).

Although these low-salinity events clearly carry elevated iron (Fig. 1b, inset), the concentration increase of <1 nM is extremely modest when compared with upwelled water. Large amounts of the dissolved (<0.4 µm) iron in rivers are present mainly as colloids. These colloids aggregate as salinity increases and settle out in estuaries, which prevents most riverine iron from reaching offshore waters¹⁵. Atmospheric deposition also seems to be a relatively minor source as offshore winds were more frequent during 1998 when dissolvable iron concentrations were lowest.

The iron source in the upwelled water does not appear to originate from diffusion of dissolved Fe from shelf sediments. Measurements of the flux of dissolved iron across the sediment–water interface on the continental shelf in Monterey Bay, using benthic flux chambers^{16,17}, average 5 ± 3 µmol Fe m⁻² d⁻¹ (standard error, $n = 11$; K.S.J., unpublished results). If the shelf has a mean depth of 50 m, then the flux of dissolved iron can add only 0.1 nM Fe d⁻¹, which is an insignificant amount.

Vertical profiles of dissolvable iron were collected along the transect during the S298 cruise to assess the amount of dissolvable iron in the subsurface waters (Fig. 3a). The concentrations of dissolvable iron at a station 200 km offshore are no higher than 2.3 nM Fe in the upper 1,000 m of the water column. Concentrations below the mixed layer at the Monterey Bay station are as high as 110 nM, in contrast to the low values seen at the surface (Fig. 3a). There is little difference in the profiles of nitrate concentration at these two stations (Fig. 3c). The high dissolvable iron concentrations in subsurface waters of Monterey Bay must be derived from resuspended sediments⁹, as light transmission is markedly lower in waters with high iron (Fig. 3b).

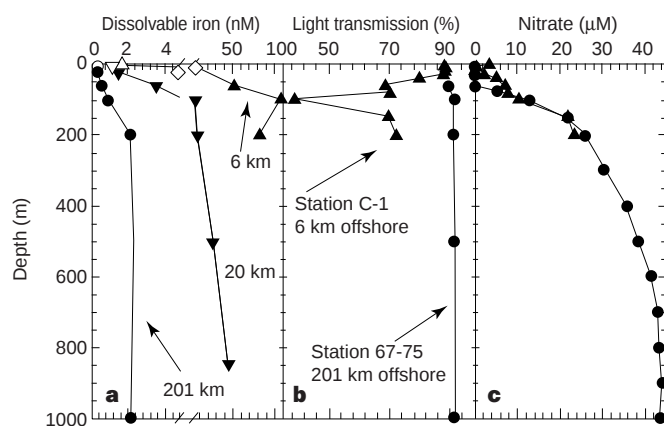


Figure 3 Vertical profiles of dissolvable iron, light transmission and nitrate. Measurements were made on cruise S298 at station 67–75 (35.95° N, 123.84° W), 201 km offshore near the offshore end of the transect shown in Fig. 1, and at station C-1 (36.80° N, 121.85° W), 6 km offshore. An iron profile is also shown from station M-1 (36.78° N, 122.01° W), 20 km offshore. Dissolvable iron was determined as described above, except that samples were acidified at pH 3 for as long as 8 h before analysis. Iron measurements made immediately using the pumping system are shown as open symbols, whereas filled symbols represent samples collected with a rosette. The iron concentrations determined on rosette samples in the mixed layer at C-1 (open diamonds) appear to be too high, perhaps owing to inadequate flushing of the rosette bottles as they were raised to the surface, rather than to the longer time at low pH, as evidenced by laboratory experiments²⁶ and the agreement of pump and surface-rosette samples at other stations. Light transmission was determined with a SeaTech 25 cm transmissometer. Nitrate was determined by colorimetry on frozen samples returned to a shore-based laboratory.

The large amounts of dissolvable iron found at the surface during upwelling conditions must be introduced by resuspension of particles into the benthic boundary layer as the upwelling source waters flow over the continental shelf. This iron-enriched water is then upwelled to the surface. These conclusions are supported by recent observations on the New Zealand shelf¹⁸. Resuspension of iron clearly occurred during the El Niño conditions of 1998, but either a lack of offshore transport trapped the iron in the inner-shelf region or weaker mixing trapped it below the surface, and surface concentrations remained low.

Upwelled water requires a large amount of iron to make use of all of the nitrate quickly. A C/Fe ratio of 27,000 was estimated from increases in particulate carbon and iron concentrations during the diatom bloom observed in the IronEx II equatorial iron fertilization experiment (K.S.J. *et al.*, manuscript in preparation). Similar C/Fe ratios are observed in cultures of coastal diatoms growing at their maximum rate¹⁹. Upwelled water with 20 μM nitrate would need at least 5 nM Fe (20 μM N × 106 μmol C/16 μmol N × 1 μmol Fe/27,000 μmol C) for diatoms to consume all of the nitrate. Dissolved (<0.4 μM) iron concentrations in the upwelled source waters are less than 1.5 nM (ref. 3), which is not sufficient to support use of the upwelled nitrate. The measurements of dissolvable iron reported here include both dissolved and labile particulate iron. A fraction of the iron in particles will be unavailable to plankton and some will be lost as particles sink from the system. These measurements must represent an upper limit on the amount of biologically available iron in the upwelled water. However, large decreases in the Fe/Al ratio in particles sampled within the CCS are strong evidence that some of this particulate iron is consumed³.

Concentrations of dissolvable iron that were sufficient for full utilization of the nitrate were observed on the S197 cruise to the coastal upwelling front. These high iron concentrations (>2 nM at the surface, up to 200 km offshore) (Fig. 1a) seem to have originated

close to shore (Fig. 3). This has important consequences as upwelling processes do not always originate in the inner-shelf region. Dynamic processes associated with the jets and eddies of the CCS⁴ or large-scale geostrophic adjustments of the thermocline²⁰ have been implicated as important for stimulating phytoplankton productivity in eastern boundaries. Coastal upwelling from the inner-shelf region, which can be important in the CCS²¹, must be the most efficient mechanism for injection of iron from the benthic boundary layer. The other processes bring macronutrients into the euphotic zone, but they may lack the continental iron source, which leads to development of high nitrate, low chlorophyll conditions. The region of the upwelling source can also migrate from the inner shelf to the shelf break as the wind strength increases in some eastern boundary current systems²², which would result in a time-varying iron source. Although high nitrate, low chlorophyll conditions have been observed in the CCS^{4,6}, their abundance and persistence have not been well resolved. When the coastal upwelling process dominates, such as during March 1997 and the spring and early summer of 1987 (ref. 4), high nitrate, low chlorophyll conditions are unlikely.

These differences in iron-source strength must contribute to the variability of the biological response to upwelling seen in coastal upwelling ecosystems throughout the world²³. The low iron concentrations observed during the S298 cruise also demonstrate that iron may be deficient at ocean margins where upwelling does not occur, even if there is a large riverine input. These conclusions indicate that iron limitation of coastal systems could have been universal phenomena during the low stands of sea level at glacial maxima, when narrow continental shelves are a global feature. □

Received 18 January; accepted 16 March 1999.

- Martin, J. H. *et al.* Testing the iron hypothesis in ecosystems of the equatorial Pacific Ocean. *Nature* **371**, 123–129 (1994).
- Coale, K. H. *et al.* A massive phytoplankton bloom induced by an ecosystem-scale iron fertilization experiment in the equatorial Pacific Ocean. *Nature* **383**, 495–501 (1996).
- Johnson, K. S., Gordon, R. M. & Coale, K. H. What controls dissolved iron in the world ocean? *Mar. Chem.* **57**, 137–161 (1997).
- Chavez, F. P. *et al.* Horizontal transport and the distribution of nutrients in the Coastal Transition Zone off Northern California: effects on primary production, phytoplankton biomass and species composition. *J. Geophys. Res.* **96**, 14833–14848 (1991).
- Hutchins, D. A. & Bruland, K. W. Iron-limited diatom growth and Si:N uptake ratios in a coastal upwelling regime. *Nature* **393**, 561–564.
- Hutchins, D. A., DiTullio, G. R., Zhang, Y. & Bruland, K. Y. An iron limitation mosaic in the California upwelling regime. *Limnol. Oceanogr.* **43**, 1037–1054 (1998).
- Chavez, F. P. & Toggweiler, J. R. in *Upwelling in the Ocean: Modern Processes and Ancient Records* (eds Summerhayes, C. P., Emeis, K.-C., Angel, M. V., Smith, R. L. & Zeitschel, B.) 313–320 (Wiley, Chichester, 1995).
- Ryther, J. H. Photosynthesis and fish production in the sea. *Science* **166**, 72–76, (1969).
- Martin, J. H. & Gordon, R. M. Northeast Pacific iron distributions in relation to phytoplankton productivity. *Deep-Sea Res.* **35**, 177–196 (1988).
- Wu, J. & Luther, G. W. Spatial and temporal distribution of iron in the surface water of the northwestern Atlantic Ocean. *Geochim. Cosmochim. Acta* **60**, 2729–2741 (1996).
- Chavez, F. P. *et al.* Moorings and drifters for real-time interdisciplinary oceanography. *J. Atmos. Oceanic Technol.* **14**, 1199–1211 (1997).
- Lynn, R. J. & Simpson, J. J. The California Current System: the seasonal variability of its physical characteristics. *J. Geophys. Res.* **92**, 12947–12966 (1987).
- Rosenfeld, L. K., Schwing, F. B., Garfield, N. & Tracy, D. E. Bifurcated flow from an upwelling center: a cold water source for Monterey Bay. *Cont. Shelf Res.* **14**, 931–964 (1994).
- Breaker, L. C. & Broenkow, W. W. The circulation of Monterey Bay and related processes. *Oceanogr. Mar. Biol. Ann. Rev.* **32**, 1–64 (1994).
- Boyle, E. A., Edmond, J. M. & Sholkovitz, E. R. The mechanism of iron removal in estuaries. *Geochim. Cosmochim. Acta* **41**, 1313–1324 (1977).
- Johnson, K. S. *et al.* Manganese flux from continental margin sediments in a transect through the oxygen minimum. *Science* **257**, 1242–1245 (1992).
- McManus, J., Berelson, W. M., Coale, K. H., Johnson, K. S. & Kilgore, T. E. Phosphorous regeneration in continental margin sediments. *Geochim. Cosmochim. Acta* **61**, 2891–2907 (1997).
- Croot, P. L. & Hunter, K. A. Trace metal distributions across the continental shelf near Otago Peninsula, New Zealand. *Mar. Chem.* **62**, 185–201 (1998).
- Sunda, W. G. & Huntsman, S. A. Iron uptake and growth limitation in oceanic and coastal phytoplankton. *Mar. Chem.* **50**, 189–206 (1995).
- Hayward, T. L. & Venrick, E. L. Near-surface pattern in the California Current: Coupling between physical and biological structure. *Deep-Sea Res.* **45**, 1617–1638 (1998).
- Lentz, S. J. Current dynamics over the northern California inner shelf. *J. Phys. Oceanogr.* **24**, 2461–2478 (1994).
- Barton, E. D., Huyer, A. & Smith, R. L. Temporal variability observed in the hydrographic region near Cabo Corveiro in the northwest African upwelling region, February to April, 1974. *Deep-Sea Res.* **24**, 7–23 (1977).
- Brink, K. H. *et al.* in *Upwelling in the Ocean: Modern Processes and Ancient Records* (eds Summerhayes, C. P., Emeis, K.-C., Angel, M. V., Smith, R. L. & Zeitschel, B.) 103–124 (Wiley, Chichester, 1995).
- Obata, H., Karatani, H. & Nakayama, E. Automated determination of iron in seawater by chelating resin concentration and chemiluminescence detection. *Anal. Chem.* **5**, 1524–1528 (1993).
- Sakamoto, C. M., Friederich, G. E., Service, S. K. & Chavez, F. P. Development of automated surface

seawater nitrate mapping systems for use in open ocean and coastal waters. *Deep-Sea Res.* **143**, 1763–1775 (1996).

26. Obata, H., Karatani, H., Matsui, M. & Nakayama, E. Fundamental studies for chemical speciation in seawater with an improved analytical method. *Mar. Chem.* **56**, 97–106 (1997).

Acknowledgements. We thank G. Elrod, E. Guenther, C. Hunter, J. Nowicki and S. Tanner for the iron analyses and assistance with sampling, and the crews of the research vessels *Western Flyer* and *New Horizon* for providing valuable assistance at sea. Funding was provided by the David and Lucile Packard Foundation through MBARI and by the National Science Foundation.

Correspondence and requests for materials should be addressed to K.S.J. (e-mail: johnson@mlml.calstate.edu).

Archaeopteris is the earliest known modern tree

Brigitte Meyer-Berthaud*, Stephen E. Scheckler† & Jobst Wendt‡

* Laboratoire de Paléobotanique, ISEM, Université Montpellier 2, Place Bataillon, F-34095 Montpellier Cedex 5, France

† Department of Biology, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061-0406, USA

‡ Universität Tübingen, Geologisch-Paläontologisches Institut, Sigwartstrasse 10, D-72076 Tübingen, Germany

Archaeopteris is an extinct plant which is of botanical interest for two reasons. It was the main component of the earliest forests until its extinction around the Devonian/Carboniferous boundary^{1–3}, and phylogenetically, it is the free-sporing taxon that shares the most characteristics with the seed plants^{4,5}. Here we describe the largest group of anatomically preserved *Archaeopteris* remains ever found, from the Famennian marine beds of south-eastern Morocco⁶, and provide the first evidence that, in terms of development and branching strategies, these 370-million-year-old plants were the earliest known modern trees. This modernization involved the evolution of four characteristics: a lateral branching syndrome similar to the axillary branching of early seed plants; adventitious latent primordia similar to those produced by living trees, which eventually develop into roots on stem cuttings; nodal

zones as important sites for the subsequent development of lateral organs; and wood anatomy strategies that minimize the mechanical stresses caused by perennial branch growth.

Archaeopteris is thought to have been an excurrent tree, with a single trunk producing helically arranged deciduous branches growing almost horizontally⁷. All studies relating to the development of *Archaeopteris* support the view that these ephemeral branches arise from the pseudomonopodial division of the trunk apex^{8–11}. Apical branching also characterizes all other contemporaneous non-seed-plant taxa including those that had also evolved an arborescent habit, such as lepidosigillarioid lycopsids and cladoxy-lalean ferns. This pattern, which can disadvantage the tree if the trunk apex is damaged, contrasts with the axillary branching reported in early seed plants¹². Analysis of a 4 m-long trunk from the Famennian of Oklahoma has shown that *Archaeopteris* may produce adventitious organs¹³. These have been interpreted as being branches of limited lifespan that increase the survival potential of the plant.

About 150 specimens were collected in three localities of the Mader Basin and Tafilalt Platform in eastern Anti-Atlas⁶. They were found in dark-grey shales with calcareous concretions of the lowermost Famennian age (Kellwasser facies, *crepida* zone). They range from 2.5 mm-wide distal axes to portions of trunks nearly 40 cm in diameter. Developmental analyses were conducted on one 5 m-long decorticated portion of trunk, which was presumed to be that of a young individual given its proximal diameter of less than 10 cm (ref. 14). Three types of appendage were identified from the size, structure, length and spatial arrangement of vascular traces along a 40 cm-long portion of its distal part (Fig. 1a). Developmentally, type A organs correspond to the ephemeral, apically produced branches previously recognized in *Archaeopteris*. We interpret them as being physiologically homologous to seed-plant leaves and assume that they were not significant in the large-scale architectural structure of the tree.

Type H adventitious organs are also short-lived. Their traces, which occur singly or in serial groups, compare well in size, structure and horizontal course with traces to small adventitious appendages in Trivett's¹³ *Archaeopteris* trunk from Oklahoma. New

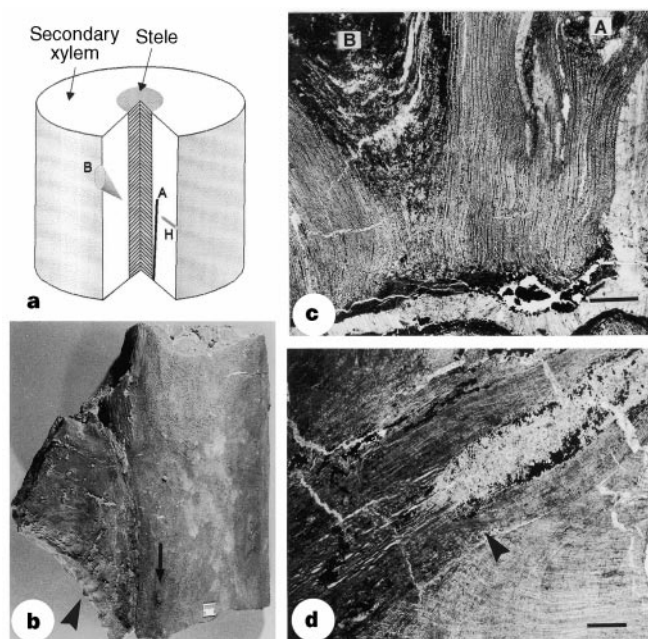


Figure 1 *Archaeopteris*, Early Famennian of Morocco. **a**, The relative arrangement of type A, B and H traces in the trunk. **b**, Trunk exterior with a large branch on the left that has a pronounced collar of trunk wood surrounding its base, showing transverse waves of wood on the branch bottom (arrowhead) and a cluster of

small buried traces on the branch flank (arrow). **c**, Cross-section of trunk showing a type B trace lateral to a type A one at the level of occlusion. **d**, Type B trace in radial section showing trace to lateral appendage (arrowhead). Scale bars: **b**, 1 cm; **c**, **d**, 1 mm.

information from our specimen reveals that they usually occur on the same radius as a type A branch and that they are similar to the radial bands of tissue that differentiate in the wood of living trees when latent primordia are produced¹⁵. These tissue bands are generally associated with the development of roots for vegetative propagation.

The type B branches of *Archaeopteris* represent a major innovation (Fig. 1c–d); five are recorded in the portion of trunk analysed. These laterals do not arise in the same ontogenetic spiral as the type A branches. They represent a new category of organ, of possibly adventitious origin, where the innermost part of its vascular trace, although close to the outer edge of the trunk primary vascular cylinder, apparently does not connect to it. Their sequence of production is irregular and currently unpredictable but the site where they originate is spatially determined, as they always occur on a radius next to the site of attachment of type A branches (Fig. 1c). Type B branches correspond to the largest traces running into the trunk, which suggests that they had a larger morphogenetic potential than the type A ones. Type B traces depart at approximately 30° from the horizontal and protrude on the outer surface of the trunk (Fig. 1d). They produce regularly arranged lateral appendages. The occurrence of short internodes in the basal part supports the view that type B organs had a delayed development. We interpret these branches as being long-lived structures that probably represent significant permanent constructional units of the tree architecture.

The earliest evidence of axillary branching was reported for Tournaisian seed plants of the family Calamopityaceae¹². Branches of *Calamopitys* are inserted in the axils of leaves but their traces connect to the closest vascular traces, not necessarily to the main stem, which suggests some period of branch primordium dormancy. The new *Archaeopteris* lateral branching syndrome, which involves the evolution of an extra type of branch and its spatial dependence on a leaf-type organ, apparently needs only a few positional adjustments to fit with the axillary branching of Calamopityaceae. It represents a major breakthrough from pseudomonopodial types of branching and may correspond to an evolutionary step towards the axillary branching of basal seed plants.

Several trunk fragments from Morocco bear the bases of large branches at wide angles. These show characteristic waves of undulating wood at their bottom points of attachment and collars of extra wood surrounding the sides and upper surface (Fig. 1b). The points of attachment of these perennial branches are reinforced by these external collars and by internal sockets for compressional and tensional load-bearing, just like modern tree limbs^{16–18} that grow heavier each year.

New information on branching patterns in *Archaeopteris* shows that this Devonian tree evolved most of the developmental features subsequently selected in the vegetative body of most derived seed plants. These features allowed the emergence of new growth forms which lived longer and could occupy space more efficiently, thereby increasing their fitness. They may partly explain the worldwide dominance of *Archaeopteris* forests on Late Devonian floodplains. □

Received 15 January; accepted 8 March 1999.

1. Fillion-Demaret, M. Some uppermost Devonian megafossils: a stratigraphical review. *Ann. Soc. Géol. Belgique* **109**, 43–48 (1986).
2. Beerbower, R. et al. in *Terrestrial Ecosystems Through Time* (ed. Behrensmeier, A. K.) 205–325 (Univ. of Chicago Press, Chicago, 1992).
3. Schekler, S. E. Geology, floristics, and paleoecology of Late Devonian coal swamps from Appalachian Laurentia (U.S.A.). *Ann. Soc. Géol. Belgique* **109**, 209–222 (1986).
4. Rothwell, G. W. & Serbet, R. Lignophyte phylogeny and the evolution of spermatophytes: a numerical cladistic analysis. *Syst. Bot.* **19**, 443–482 (1994).
5. Nixon, K. C., Crepet, W. L., Stevenson, D. & Friis, E. M. A reevaluation of seed plant phylogeny. *Ann. Missouri Bot. Gard.* **81**, 484–533 (1994).
6. Wendt, J. & Belka, Z. Age and depositional environment of Upper Devonian (Early Frasnian to Early Famennian) black shales and limestones (Kellwasser facies) in the eastern-Anti-Atlas, Morocco. *Facies* **25**, 51–90 (1991).
7. Beck, C. B. & Wight, D. C. in *Origin and Evolution of Gymnosperms* (ed. Beck, C. B.) 1–84 (Columbia University Press, New York, 1988).
8. Beck, C. B. On the anatomy and morphology of lateral branch systems of *Archaeopteris*. *Am. J. Bot.* **58**, 758–784 (1971).

9. Carluccio, L. M., Hueber, F. M. & Banks, H. P. *Archaeopteris macilentia*, anatomy and morphology of its front. *Am. J. Bot.* **53**, 719–730 (1966).
10. Kenrick, P. & Fillion-Demaret, M. *Archaeopteris roemeriana* (Göppert) sensu Stockmans, 1948 from the Upper Famennian of Belgium: anatomy and leaf polymorphism. *Bull. Inst. R. Sci. Nat. Belgique, Sci. Terre* **61**, 179–195 (1991).
11. Schekler, S. E. Ontogeny of progymnosperms. II. Shoots of Upper Devonian Archaeopteridales. *Can. J. Bot.* **56**, 3136–3170 (1978).
12. Galtier, J. & Holmes, J. New observations on the branching of Carboniferous ferns and pteridosperms. *Ann. Bot.* **49**, 737–746 (1982).
13. Trivett, M. L. An architectural analysis of *Archaeopteris*, a fossil tree with pseudomonopodial and opportunistic adventitious growth. *Bot. J. Linn. Soc.* **111**, 301–329 (1993).
14. Meyer-Berthaud, B., Wendt, J. & Galtier, J. First record of a large *Callixylon* trunk from the Late Devonian of Gondwana. *Geol. Mag.* **134**, 847–853 (1997).
15. Fink, S. Adventitious root primordia—The cause of abnormally broad xylem rays in hardwood and softwoods. *LAWA Bull. n.s.* **3**, 31–38 (1982).
16. Larson, P. R. *The Vascular Cambium: Development and Structure* 1–725 (Springer, Berlin, 1994).
17. Mattheck, C. & Kubler, H. *Wood—The Internal Optimization of Trees* 1–129 (Springer, Berlin, 1995).
18. Zobel, B. J. & van Buijtenen, J. P. *Wood Variation: its Causes and Control* 1–363 (Springer, Berlin, 1989).

Acknowledgements. Portions of this study were supported by a sabbatical research leave from VPI & SU and a grant from the National Science Foundation (USA), Division of Integrative Biology and Neuroscience, Program for Ecological and Evolutionary Physiology (to S.E.S.). We thank M. Boutaleb and A. Fadile for the issue of working permits and the export of samples; S. Doering, R. Feist, B. Kaufmann, C. Klug and D. Korn for advice in the field; and J. Guiraud and L. Meslin for technical help. This work was supported by grants from IGCN (North Gondwanan mid-Palaeozoic biodynamics) and CNRS.

Correspondence and requests for materials should be sent to B.M.-B. (e-mail: meyerber@isem.univ-montp2.fr).

Fungus-growing ants use antibiotic-producing bacteria to control garden parasites

Cameron R. Currie^{*†}, James A. Scott[†],
Richard C. Summerbell^{‡§} & David Malloch[†]

^{*} Smithsonian Tropical Research Institute, PO Box 2072, Balboa, Republic of Panama

Departments of [†] Botany and [‡] Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario M5S 3B2, Canada

[§] Ontario Ministry of Health, 81 Resources Road, Toronto, Ontario M9P 3T1, Canada

The well-studied, ancient and highly evolved mutualism between fungus-growing ants and their fungi has become a model system in the study of symbiosis^{1–5}. Although it is thought at present to involve only two symbionts, associated with each other in near isolation from other organisms^{1–5}, the fungal gardens of attine ants are in fact host to a specialized and virulent parasitic fungus of the genus *Escovopsis* (Ascomycotina)⁶. Because the ants and their fungi are mutually dependent, the maintenance of stable fungal monocultures in the presence of weeds or parasites is critical to the survival of both organisms. Here we describe a new, third mutualist in this symbiosis, a filamentous bacterium (actinomycete) of the genus *Streptomyces* that produces antibiotics specifically targeted to suppress the growth of the specialized garden-parasite *Escovopsis*. This third mutualist is associated with all species of fungus-growing ants studied, is carried upon regions of the ants' cuticle that are genus specific, is transmitted vertically (from parent to offspring colonies), and has the capacity to promote the growth of the fungal mutualist, indicating that the association of *Streptomyces* with attine ants is both highly evolved and of ancient origin.

Because few organisms cultivate their own food, fungus-gardening by ants (Attini: Formicidae) is considered to be a major breakthrough in animal evolution⁷. These ants forage on a variety of substrates that they use for the cultivation of the vegetative mycelium of a fungus, their dominant food source. Fungus cultivation evolved apparently only once in the attines, over 50 million years ago, with the domestication of a fungus in the family Lepiotaceae (Agaricales: Basidiomycotina)^{3,4,8}. Other lepiotaceous lineages, and in one case a distantly related non-lepiotaceous basidiomycete, were

domesticated in subsequent evolutionary history⁴. The success of fungal cultivation by the attine ants is illustrated by the leaf-cutting genera, *Acromyrmex* and *Atta*, which are the dominant herbivores in the neotropics⁹.

Certain areas of the cuticle of fungus-growing ants are coated with what appears to the naked eye to be a powdery, whitish-grey crust (Fig. 1). This has been dismissed previously as a 'waxy bloom', implying that its aetiology was cuticular exudate¹⁰. Micromorphological and biochemical examination reveals that this coating is in fact formed from masses of a filamentous bacterium (actinomycete) of the genus *Streptomyces* (Fig. 2a; see Methods). Actinomycetes are mostly soil-dwelling organisms of great abundance and ecological importance that produce an array of secondary metabolites, many of which have specific antibacterial or antifungal properties^{11,12}. In fact, most antibiotics developed for human pharmaceutical use are actinomycete metabolites, many derived from the genus *Streptomyces*^{11,12}. In light of the unique biochemical properties of actinomycetes as a group, we proposed that the *Streptomyces* associated with fungus-growing ants may have an important function in this symbiosis, that of suppressing the growth of potentially devastating pathogens.

To ascertain the distribution and prevalence of this bacterium, both within and between species and genera of fungus-growing ants, we studied 22 species of attine ants representing 8 genera for the presence of the actinomycete. The bacterium was associated with all species studied, from the phylogenetically basal genera *Myrmecocrypta* and *Apterostigma* to the highly derived, leaf-cutting genera *Acromyrmex* and *Atta*. All 112 colonies from Panama sampled for the presence of the actinomycete in 1997 and 1998 showed this bacterial association. In all cases, the actinomycete was concentrated on genus-specific areas of the ant integument that appear to be modified for the maintenance and growth of the *Streptomyces*, conceivably to facilitate the distribution of bacterial metabolites throughout the garden (Fig. 2, Table 1). In workers and queens of *Myrmecocrypta* and *Apterostigma*, the actinomycete mutualist occurs under the forelegs (Fig. 2b); in the more phylogenetically derived genera (Table 1), the actinomycete mutualist is most prominent on the laterocervical plates of the propleura (Fig. 2c), ventral, collar-like structures immediately posterior to the mouth-parts. The consistent association of the actinomycete with phylogenetically diverse attine ants, as well as its location on the ant integument (that is, conserved within species but unique

between genera), indicates that this association is highly evolved and may be of ancient origin.

We studied the presence of the actinomycete on foundress queens (gynes) during their mating flights to determine whether, like the fungal mutualist, this bacterium is transmitted vertically between parent and offspring colonies. We examined 74 foundress queens and 15 males collected from throughout Gamboa, Panama, during the mating flight of *Acromyrmex octospinosus* on 9 May 1997. *Streptomyces* was present on the cuticle of all gynes examined, whereas it was absent from all males collected during the same mating flight. Examination of female and male reproductive individuals in natal gardens of *Trachymyrmex* cf. *zeteki* before mating flights revealed the presence of the cuticular actinomycete on females only ($n = 74$ ants, including 43 males, from 10 colonies). Because males do not participate in the founding of new colonies or



Figure 1 Photograph showing the presence of the third mutualist, *Streptomyces*, on the cuticle of *Acromyrmex octospinosus*.

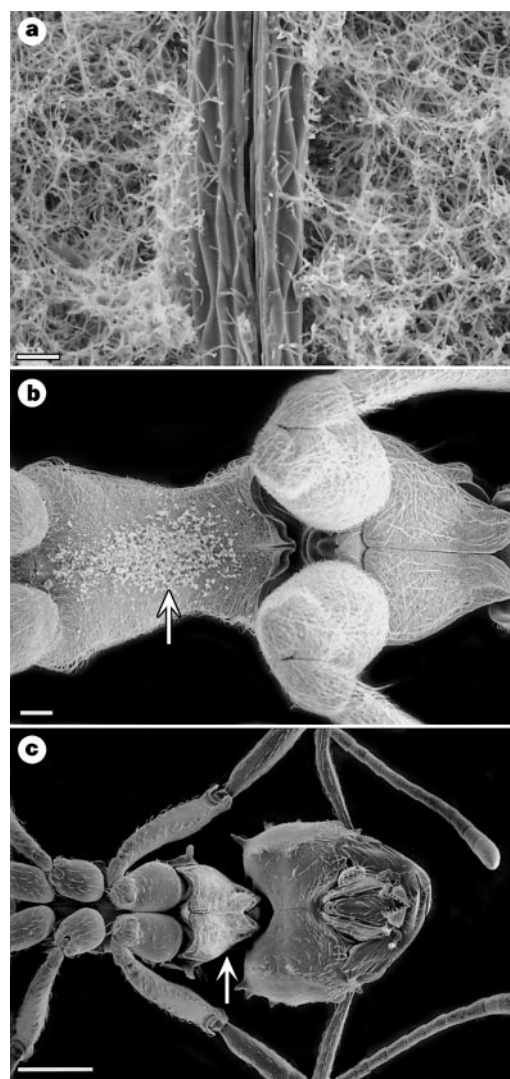


Figure 2 Scanning electron micrographs of fungus-growing ants, showing the location of *Streptomyces*. **a**, View of the filamentous growth form of the actinomycetous bacterium, showing the typical growth pattern and its thickness on the cuticle. Scale bar represents 10 μm . **b**, View of *Streptomyces* (arrow) under the forelegs of *Apterostigma* sp; this is the characteristic location of the bacterium in phylogenetically basal genera of the Attini. Scale bar represents 100 μm . **c**, Ventral view of a minimum worker of *Acromyrmex octospinosus*. The actinomycete-laden laterocervical plate (arrow) can be seen just below the head on the propleura of the ant. This is the characteristic location of the bacterium on the more phylogenetically derived genera of the Attini. Scale bar represents 500 μm .

Table 1 Location of *Streptomyces* on different genera of attine ants

Attine genera*	Under forelegs on propleura	Laterocervical plates of propleura
<i>Myrmicocrypta</i>	Yes	
<i>Apterostigma</i>	Yes	
<i>Mycocepurus</i>	Yes	Yes
<i>Cyphomyrmex</i>		Yes
<i>Trachymyrmex</i>		Yes
<i>Acromyrmex</i>		Yes
<i>Atta</i> †	Not visible on exoskeleton	Not visible on exoskeleton

* Attine genera listed from most phylogenetically basal to most phylogenetically derived^{8,23}.

† Although not visible on exoskeleton, *Streptomyces* was isolated repeatedly from intact *Atta* spp.

in tending the fungal garden, these data support the proposed role of the actinomycete in suppressing the growth of garden pathogens.

To determine whether the attine-associated actinomycete produces compounds with beneficial antimicrobial properties, we performed bioassays in which taxonomically diverse sets of fungi were challenged with attine *Streptomyces* isolates. *Streptomyces* isolates obtained from *Acromyrmex octospinosus* lacked detectable inhibitory effects on the growth of generalist saprotrophic fungi, entomopathogenic fungi, and other fungi commonly used for antibiotic screening. However, it showed potent inhibitory effects towards *Escovopsis*, a fungal genus (anamorphic Hypocreales: Ascomycotina)¹³ identified recently as a specialized virulent parasite of the attine fungal gardens⁶. The actinomycete completely suppressed spore germination of *Escovopsis* isolates in 25% of bioassays. Linear fungal growth in the remaining challenges was inhibited by $73.9 \pm 3.0\%$ (mean $\pm$ s.e.m.), typically resulting in zones of inhibition larger than 30 mm (Fig. 3). Other bioassay challenges between the corresponding *Streptomyces* and *Escovopsis* from colonies of *Cyphomyrmex longiscapus*, *Atta colombica* and *Atta cephalotes* again resulted in significant inhibition of the growth of *Escovopsis*.

We tested growth-enhancing effects of the actinomycete on the basidiomycete mutualist in broth culture bioassays, using a *Streptomyces* isolate obtained from the phylogenetically basal attine genus *Apterostigma*. We observed significant increases in basidiomycete vegetative biomass in the presence of the actinomycete culture filtrate (vegetative biomass in presence of actinomycete culture filtrate averaged 47.9 ± 7.6 mg dry weight, versus 5.3 ± 2.4 mg dry weight for unamended controls; significant at $P = 0.0029$, Student's *t*-test). This increase in biomass may result from the production of growth-promoting compounds by the actinomycete (for example, vitamins, enzymes, and/or amino acids)^{14–16}.



Figure 3 Bioassay challenge between *Streptomyces* and *Escovopsis*, the specialized parasite of attine fungal gardens, associated with *Acromyrmex octospinosus*, illustrating the substantial zone of inhibition of fungal growth.

Several lines of evidence indicate that this newly discovered bacterial symbiont of the attine ants is a third mutualist in an ancient symbiosis among the ants, the domesticated fungi, the parasitic fungi, and the antibiotic-producing bacteria. First, this actinomycete is associated with all attine species and all colonies studied, representing the generic diversity of the Attini. Second, the actinomycete is transmitted vertically from parent to daughter nest, as is the fungal mutualist, and third, it promotes the growth of the fungal mutualist *in vitro*. Fourth, and most important, this *Streptomyces* produces highly potent antibiotics that selectively inhibit the growth of the garden parasite *Escovopsis*. Although many antibiotics developed for medicinal use have fairly broad-spectrum effects, it is likely that most fungicidal secondary metabolites produced by microbes evolved toward specific targets, such as competitors and pathogens^{17,18}. As the production of secondary metabolites is energetically costly and requires complex, genetically based biosynthetic pathways, its evolution and subsequent maintenance is presumed to impart a substantial selective advantage to the microbe¹⁷. Thus the production of antibiotics by the attine-associated *Streptomyces* that specifically target the growth of *Escovopsis* is compelling evidence that the bacterium is a highly evolved mutualist. The ants use this bacterium, because of its production of antibiotics, to treat microbial infection of their garden, and, in exchange, the ants disperse the actinomycete and appear to provide some form of nourishment for its growth.

Although the ant–fungus mutualism is often regarded as one of the most fascinating examples of a highly evolved symbiosis, it is now clear that its complexity has been greatly underestimated. The attine symbiosis appears to be a co-evolutionary ‘arms race’ between the garden parasite, *Escovopsis*, on the one hand, and the tripartite association amongst the actinomycete, the ant hosts, and the fungal mutualist on the other. The evolution of a mutualistic association between the attine ants and an actinomycete that suppresses parasites is perhaps not surprising. The benefits of such a symbiosis are illustrated by the paramount part played by therapeutic antibiotics in human biomedical history. Our results indicate that microbes and their metabolites may be key components in the regulation of other symbiotic associations between higher organisms, and thus a more detailed analysis of their functions promises to illuminate the general dynamics of symbiosis. Study of the presumably highly evolved chemical interactions between symbionts may provide valuable theoretical and practical insights regarding the identification, production and application of antibiotics^{19–21}. □

Methods

Identification of bacterium. To identify the bacterium, we used micro-morphological parameters as well as accepted biochemical criteria. We analysed the cell-wall fatty-acid methyl esters by gas–liquid chromatography²². The absence of tuberculosteric acid and related 10-methyl esters excluded the genera *Actinoadura* and *Nocardopsis*, which are morphologically similar to *Streptomyces*.

Attines. The presence of actinomycete was determined in fungus-growing ant species from the canal region of Panama (112 colonies, 17 species) and the Napo province of Ecuador (25 colonies, 5 species). Attine genera studied included a representative sampling of both ‘lower’ (phylogenetically basal) and ‘higher’ (phylogenetically derived) attines, namely *Myrmicocrypta* (two species), *Apterostigma* (four species), *Mycocepurus* (two species), *Cyphomyrmex* (two species), *Sericomyrmex* (one species), *Trachymyrmex* (five species), *Acromyrmex* (three species) and *Atta* (three species).

Antibiotic-bioassay challenges. Bioassays were done in Petri dishes using an actinomycete isolate obtained from a phylogenetically derived attine species, *Acromyrmex octospinosus*. Saprotrophic fungi isolated from attine gardens, including taxa closely related to *Escovopsis*, were tested, as were generalist entomopathogens. We also assayed a diverse, representative set of fungi used for general anti-fungal screening. Finally, we challenged representative strains of the specialist garden parasite *Escovopsis*. Specifically, two strains of *Metarhizium anisopliae* and one strain of *Beauveria bassiana* were challenged

with actinomycete, as were the following common microfungi: *Absidia* sp., *Ascobolus crenulatus*, *Aspergillus fumigatus*, *Coprinus patouillardii*, *Cryptococcus albidus*, *Drechslera biseptata*, *Exophiala spinifera*, *Fusarium oxysporum*, *Mucor pyriformis*, *Penicillium* sp., *Pythium aphanidermatum*, *Schizophyllum commune*, *Sordaria fimicola* and *Trichoderma* sp.

Each *Streptomyces*–fungal challenge was replicated three times and done on Czapek yeast autolysate agar. The actinomycete was inoculated on Petri dishes and grown to a diameter of ~1.5 cm; fungal strains were then point-inoculated near the edge of the culture. Challenges were monitored every two days and growth inhibition of tested fungi was scored as a reduction of growth rate as compared with growth of fungal cultures in the absence of the *Streptomyces*, or as complete suppression of growth. We assayed possible antibiotic production specific to the specialized parasite *Escovopsis* in the same way that we assayed antibiotic production specific to other potential contaminants, except that each challenge to *Escovopsis* was replicated five times. Four strains of *Escovopsis* isolated from the gardens of different *Acromyrmex octospinosus* colonies in Panama in 1997 were tested against *Streptomyces*. We also studied the production of antibiotics specific towards *Escovopsis* in other attine species, including *Cyphomyrmex longiscapus*, *Atta colombica* and *Atta cephalotes*. The presence of a zone of inhibition in bioassays indicates first, the production of diffusible metabolites by the actinomycete, and second, the susceptibility of the test fungus to these compounds. As inhibition is dose dependent, the detection of partial inhibition implies the existence of a dose that could impart complete inhibition.

Growth-promotion bioassays. Broth cultures of the attine fungus isolated from an *Apterostigma* colony were grown with extracts from the *Streptomyces* isolated from this species. Actinomycete extracts were obtained by growing *Streptomyces* in Czapek yeast autolysate broth for 2 weeks and then passing the broth through a low protein-binding, sterilizing filter unit (Millipore, Millex) to remove bacterial biomass. We replicated each bioassay five times and used 50 ml Czapek yeast autolysate broth per bioassay.

Received 20 November 1998; accepted 18 February 1999.

- Weber, N. The fungus growing ants. *Science* **121**, 587–604 (1966).
- Wilson, E. O. *The Insect Societies* (Belknap, Cambridge, Massachusetts, 1971).
- Chapela, I. H., Rehner, S. A., Schultz, T. R. & Mueller, U. G. Evolutionary history of the symbiosis between fungus-growing ants and their fungi. *Science* **266**, 1691–1694 (1994).
- Mueller, U. G., Rehner, S. A. & Schultz, T. R. The evolution of agriculture in ants. *Science* **281**, 2034–2038 (1998).
- North, R. D., Jackson, C. W. & Howse, P. E. Evolutionary aspects of ant–fungus interactions in leaf-cutting ants. *Trends Ecol. Evol.* **12**, 386–389 (1997).
- Currie, C. R., Mueller, U. G. & Malloch, D. The agricultural pathology of ant fungal gardens. *Proc. Natl Acad. Sci. USA* (submitted).
- Wilson, E. O. in *Fire Ants and Leaf-cutting Ants*. (Westview, Boulder, 1986).
- Schultz, T. R. & Meier, R. A phylogenetic analysis of the fungus-growing ants (Hymenoptera: Formicidae: Attini) based on morphological characters on the larvae. *Syst. Entomol.* **20**, 337–370 (1995).
- Hölldobler, B. & Wilson, E. O. *The Ants* (Belknap, Cambridge, Massachusetts, 1990).
- Weber, N. A. *Gardening Ants: The Attines* (Am. Phil. Soc., Philadelphia, 1972).
- Waksman, S. A. & Lechevalier, H. A. 1962. *The Actinomycetes*, Vol. III. *Antibiotics of Actinomycetes* (Williams & Wilkins, Baltimore, 1962).
- Goodfellow, M. & Cross, T. *The Biology of Actinomycetes* (Academic, London, 1984).
- Seifert, K. A., Samson, R. A. & Chapela, I. H. *Escovopsis aspergilloides*, a rediscovered hyphomycete from leaf-cutting ant nests. *Mycologia* **87**, 407–413 (1995).
- Martin, M. M. & Martin, J. S. The biochemical basis for the symbiosis between the ant, *Atta colombica* tompsoni, and its food fungus. *J. Insect Physiol.* **16**, 109–119 (1970).
- Hervey, A., Rogerson, C. T. & Leong, I. Studies on fungi cultivated by ants. *Brittonia* **29**, 226–236 (1978).
- Cazin, J. Jr, Wiemer, D. F. & Howard, J. J. Isolation, growth characteristics, and long-term storage of fungi cultivated by attine ants. *Appl. Env. Microbiol.* **55**, 1346–1350 (1989).
- Vining, L. C. Functions of secondary metabolites. *Annu. Rev. Microbiol.* **44**, 395–427 (1990).
- Griffin, D. H. *Fungal Physiology* (Wiley-Liss, New York, 1994).
- Eisner, T. Prospecting for nature's chemicals. *Iss. Sci. Tech.* **6**, 31–34 (1990).
- Beattie, A. J. Discovering new biological resources—chance or reason. *Bioscience* **42**, 290–292 (1992).
- Caporale, L. H. Chemical ecology: a view from the pharmaceutical industry. *Proc. Natl Acad. Sci. USA* **92**, 75–82 (1995).
- Holt, J. G. et al. (eds) *Bergey's Manual of Determinative Microbiology* 9th edn. (Williams & Wilkins, Baltimore, 1994).
- Wetterer, J. K., Schultz, T. R. & Meier, R. Phylogeny of fungus-growing ants (tribe Attini) based on mtDNA sequence and morphology. *Mol. Phylogenet. Evol.* **9**, 42–47 (1998).

Acknowledgements. This work was supported by Smithsonian and NSERC predoctoral awards (to C.R.C.) and an NSERC grant (to D.M.). C.R.C. thanks the Smithsonian Tropical Research Institute and ANAM of the Republic of Panama for assisting with the research and granting collecting permits, and U. G. Mueller for guidance, support and encouragement. We thank I. Ahmad, N. Alasti-Faridani, G. de Alba, S. Barrett, E. Bermingham, A. Caballero, J. Ceballos, S. Dalla Rosa, L. Ketch, M. Leone, G. Maggiori, S. Rand and M. Witkowska for logistical support; C. Ziegler for the photograph in Fig. 1; and K. Boomsma, J. Bot, R. Cocroft, G. Currie, J. Gloer, A. Herre, H. Herz, S. Rehner, T. Schultz, N. Straus, B. Wcislo and B. Wong for comments on this study and/or manuscript.

Correspondence and requests for materials should be addressed to C.R.C. (e-mail: currie@botany.utoronto.ca).

Relative reward preference in primate orbitofrontal cortex

Léon Tremblay* & Wolfram Schultz

Institute of Physiology and Program in Neuroscience, University of Fribourg, CH-1700 Fribourg, Switzerland

* Present address: INSERM Unit 289, Hôpital de la Salpêtrière, 47 Boulevard de l'Hôpital, F-75651 Paris, France.

The orbital part of prefrontal cortex appears to be crucially involved in the motivational control of goal-directed behaviour^{1,2}. Patients with lesions of orbitofrontal cortex show impairments in making decisions about the expected outcome of actions³. Monkeys with orbitofrontal lesions respond abnormally to changes in reward expectations^{4,5} and show altered reward preferences⁶. As rewards constitute basic goals of behaviour⁷, we investigated here how neurons in the orbitofrontal cortex of monkeys process information about liquid and food rewards in a typical frontal task, spatial delayed responding⁸. The activity of orbitofrontal neurons increases in response to reward-predicting signals, during the expectation of rewards, and after the receipt of rewards. Neurons discriminate between different rewards, mainly irrespective of the spatial and visual features of reward-predicting stimuli and behavioural reactions. Most reward discriminations reflect the animals' relative preference among the available rewards, as expressed by their choice behaviour, rather than physical reward properties. Thus, neurons in the orbitofrontal cortex appear to process the motivational value of reward-induced outcomes of voluntary action.

Neurophysiological studies of behaving monkeys and rats show that neurons in the six-layered parts of orbitofrontal cortex process motivating events, discriminate between appetitive and aversive conditioned stimuli⁹ and are active during the expectation of outcomes¹⁰. Some orbitofrontal neurons may code the value of reward objects in losing their responses when animals become satiated on particular food items¹¹. Neurons in more caudal, three- and four-layered orbitofrontal regions process gustatory and olfactory information^{12–14}.

We investigated the motivational properties of orbitofrontal neurons in macaque monkeys during a spatial delayed-response task (Fig. 1). The position of a briefly presented instruction picture indicated the target of an arm movement, and its visual features predicted specifically which of two liquid or food rewards would be delivered for correct performance at the end of the trial. A subsequent uniform trigger stimulus provoked the movement to the remembered target. The reward was delivered after a brief delay during which the animal could expect the reward. Reaction times

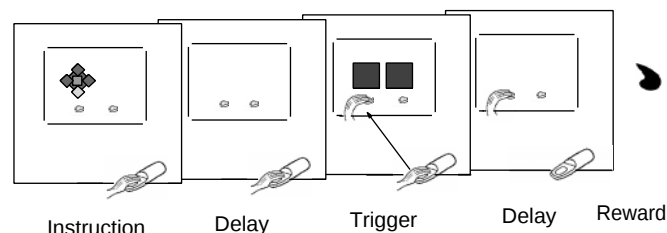


Figure 1 Spatial delayed-response task. An initial instruction picture indicates the left or right target of movement and the liquid or food reward that will be delivered at the end of the trial. Following a brief delay, two identical squares appear and the monkey moves its hand from the resting key to the left or right target lever indicated by the instruction. Correct performance is rewarded after a brief delay with a drop of liquid or a morsel of food.

differed significantly between trials with different rewards, indicating that monkeys had established reward expectations^{15,16}.

Neurons in six-layered parts of orbitofrontal areas 11 and 14 and rostral area 13 (Fig. 2) showed three principal types of activation, namely responses to instructions (15% of 1,095 tested neurons in liquid-rewarded trials, 15% of 329 neurons in food-rewarded trials), responses following reward (8% in liquid, 18% in food trials), and sustained activations preceding reward (9% in liquid, 19% in food trials). The pre-reward activations began several seconds before the reward, increased slowly and subsided <1 s after reward delivery, apparently reflecting the upcoming reward rather than preceding events. They were unrelated to arm movements, as they also occurred in non-movement trials of a delayed go/no-go task with similar temporal structure (87% of 47 neurons; L.T. and W.S., manuscript in preparation). Similar pre-reward activations occur in basal ganglia neurons^{16–18} and probably reflect the expectation of reward. In contrast, statistically significant sustained activations during the entire instruction–trigger delay were only found in two neurons, although they are frequent in dorsolateral prefrontal cortex^{19–21}. Thus, most orbitofrontal neurons modulated here were activated in response to reward-related events.

All three principal types of orbitofrontal activation discriminated between different liquid rewards and between different food rewards (Fig. 3). Activations occurred only, or were significantly

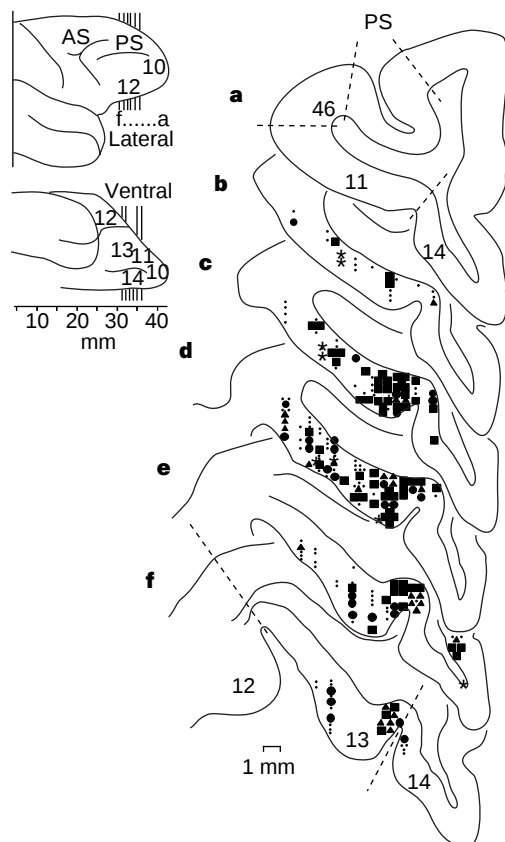


Figure 2 Positions of orbitofrontal neurons activated in liquid-reward trials. Vertical lines in the left-hand diagrams indicate rostrocaudal levels of electrode penetrations corresponding to frontal sections a–f of the right-hand diagram. Right, each symbol shows the position of a single neuron. Filled squares, neurons responding to instructions; filled circles, neurons activated pre-reward or post-reward; filled triangles, neurons responding to instructions and activated pre- or post-reward; stars, neurons with spatial relationships; small dots, activations unselective for reward, spatial position or instruction pictures. Numbers refer to architectonic areas. Area 12 was not studied. AS, arcuate sulcus; PS, principal sulcus.

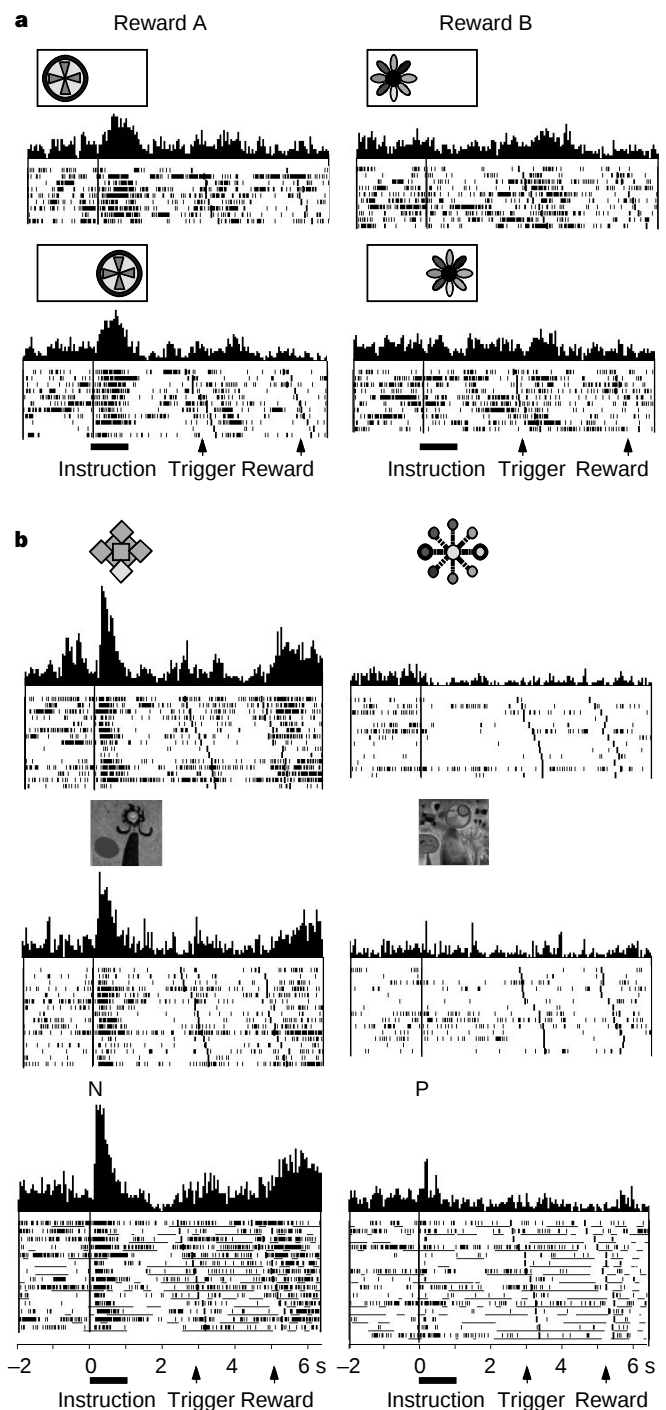


Figure 3 Reward rather than spatial or object processing in orbitofrontal neurons.

a, The response to instructions shown by a single orbitofrontal neuron that discriminates between liquid rewards but not between left and right instruction positions. The insets above the histograms show the instruction pictures used, at left or right positions. Reward A, grenadine juice; reward B, apple juice. Raster dots denote neuronal impulses aligned to instruction onset. Each line shows one trial. Use of the two rewards and positions were alternated randomly between trials; trials are separated for analysis. The original trial sequence is shown from top to bottom. **b**, Response to instructions shown by a single orbitofrontal neuron that discriminates between liquid rewards but not between visual instruction features. Instruction responses differed insignificantly between the three different instructions predicting reward A (grape juice). Reward B was orange juice. The three instruction sets were tested in separate trial blocks and are shown above the corresponding neuronal data. Thin horizontal lines in the bottom blocks indicate licking periods. This neuron also responded following delivery of reward A. N, P represent the pictures used.

higher, for one reward compared with the other rewards in 69% of 218 instruction responses, 52% of 146 reward responses and 41% of 160 pre-reward activations. These differences were unrelated to eye or licking movements before or after reward delivery; these movements varied inconspicuously between the two rewards referred to in Fig. 3b, bottom. Only 3% of 218 instruction responses discriminated between left and right targets of movement, and none of the 50 neurons tested with several instruction sets showed significantly different responses to different instructions indicating the same reward. Thus orbitofrontal neurons reflected the predicted rewards much better than they reflected the spatial or visual features of the instructions in the present task situation.

When subjects have a choice between different rewards, they may select some rewards more frequently than others. However, the preferences expressed by overt choice behaviour are relative and depend on the available alternatives. A reward that is chosen more frequently in the presence of some rewards may be instantly neglected when other, more appetizing rewards become available. Apparently, motivational values, unlike physical properties, are not fixed to individual rewards. The question arose to what extent the prominent reward relationships of orbitofrontal neurons might reflect the relative motivational values of rewards expressed by choice behaviour. This relates to the observation that monkeys with ventromedial prefrontal lesions show abnormal reward preferences⁶.

We obtained behavioural evidence for relative reward preference in a choice version of the task used above. Two instruction pictures, instead of one, were shown simultaneously above the left and right levers, respectively. Each picture was associated with a different reward. Following the trigger stimulus, animals chose one of the rewards by touching the corresponding lever. We used three food rewards (A–C) but presented only two of them in a given trial block (A and B, B and C, or A and C). Animals showed clear preferences in every comparison, choosing rewards A over B, B over C, and A over

C in 90–100% of trials, independently of the side of the touched lever. Thus reward B was chosen less frequently when reward A was available, but more frequently when reward C was available. The preference for reward B seemed to be relative and depended on the reward with which it was being compared.

We studied orbitofrontal neurons during the performance in the standard delay task described above, in which animals were presented with a single instruction picture for a single, preferred or non-preferred, reward. Two rewards alternated randomly in each trial block, and usually all three combinations of reward pairs were tested (A and B, B and C, and A and C in separate blocks). The relative preference of the animal was checked in separate choice trials before or after each recording block. The activity of 40 of 65 reward-discriminating orbitofrontal neurons indeed depended on the combination in which a particular reward occurred. The neuron of Fig. 4 showed significantly higher activation preceding reward A (preferred) as compared with reward B (non-preferred) (Fig. 4, top). When reward B was compared with reward C in a different trial block (Fig. 4, bottom), the same neuron was activated significantly more preceding reward B (now preferred) than reward C (non-preferred). Neuronal activations preceding reward B occurred only when this reward was the preferred one (Fig. 4, centre, compare top and bottom panels). Thus, this orbitofrontal neuron was activated by the reward that was relatively more preferred than the available alternative. Comparable results were seen with reward-predicting instruction responses and with responses following the reward (total of 27 neurons). The reverse was observed with 13 of the 40 neurons, which showed higher activations in response to whichever reward was less preferred than the available alternative in the A and B, B and C scheme. Neurons with activity reflecting relatively higher or lower reward preference were located in post-eriomedial area 11, close to area 14 (sections c–e of Fig. 2). Thus, the reward discrimination occurred in some orbitofrontal neurons on the basis of relative preference rather than physical properties.

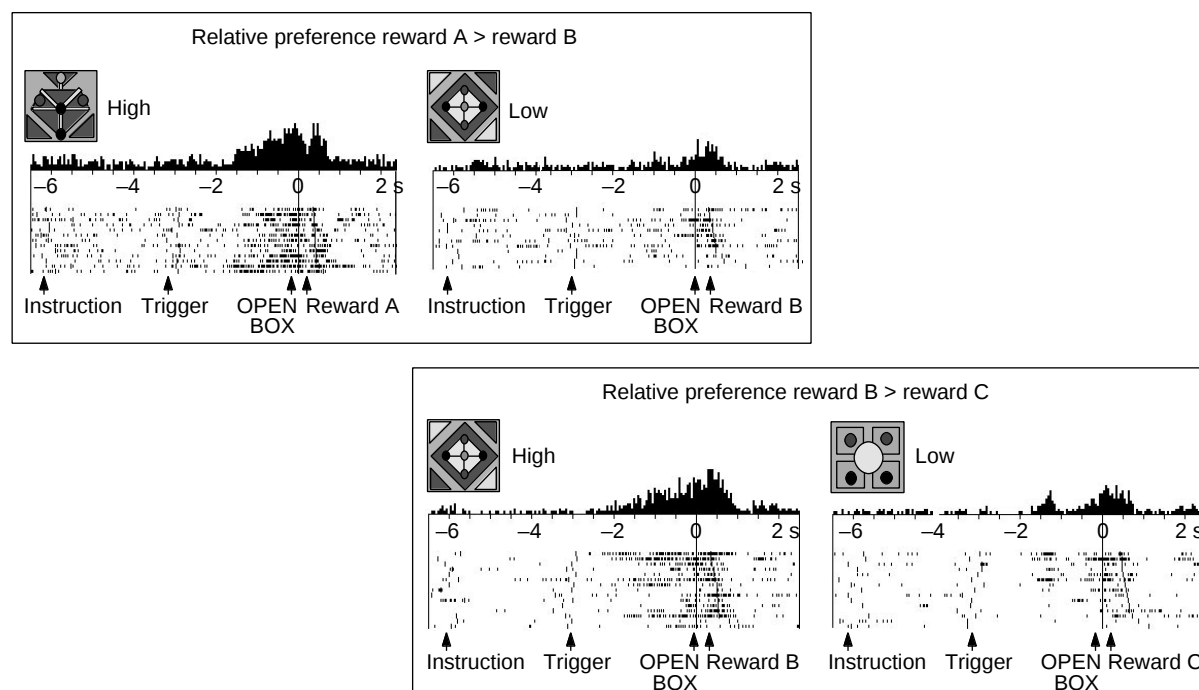


Figure 4 Coding of relative reward preference in an orbitofrontal neuron. Neuronal activity increased during the expectation period preceding the relatively more preferred food reward (raisin in top panel, apple in bottom panel). Although reward B (apple) was physically identical in the top and bottom panels, its motivational value was relative and differed according to the other available reward (its motivational value was low in the top panel and high in the bottom

panel), as expressed by the animal's choice behaviour assessed in separate trials. Different combinations of reward were used in the two trial blocks (A, raisin; B, apple; C, cereal). Each reward was specifically predicted by an instruction picture, which is shown above the histograms. Lever touch following the movement-trigger stimulus induced, 2 s later, the opening of a food box from which the animal collected the reward.

Nearly all task-related activations in the remaining 25 of the 65 neurons occurred in response to either the most preferred (A) or the least preferred (C) of the three food rewards. These activations were independent of the combination of rewards. Only a single neuron responded to the instruction for the intermediately preferred reward, B, in all comparisons.

These results show that some orbitofrontal neurons process specific aspects of motivational information. They discriminate well between different rewards, and many discriminations appear to be based on the relative preference for different rewards exhibited by animals in overt choice behaviour. The activity of these orbitofrontal neurons does not appear to code the fixed physical properties of rewards, but rather reflects the motivational value of one reward relative to another, as expressed by the behavioural preference. Just as each reward can have a higher or lower motivational value relative to the reward with which it is compared, orbitofrontal neurons can be more or less activated by one reward, depending on which alternative reward is available.

Primates are able to consume hundreds of different reward objects, but only a few objects are available at any given time. The coding of relative preferences among the available rewards would allow orbitofrontal neurons to specify motivational goals of behaviour in far more situations than would be allowed by coding of the physical properties of individual rewards.

Many neurons in both the dorsolateral and ventrolateral prefrontal cortex process spatial positions and/or visual features of environmental objects^{20–24}. These functions may be derived from inputs from posterior parietal cortex and inferotemporal cortex^{25,26}. The orbitofrontal neurons studied here seem to process basic motivational aspects of environmental events that determine the probability and intensity of goal-directed behaviour. These activities may result from trans-synaptic inputs from the striatum, which shows prominent relationships to reward expectation^{16–18}, and from inputs from the amygdala and rostral and medial temporal lobe^{27,28}. Interestingly, the delivery of reward also produces responses in ventrolateral prefrontal neurons²⁹, and expected rewards influence spatial delay activity of dorsolateral prefrontal neurons³⁰. □

Methods

Behavioural task. Two *Macaca fascicularis* monkeys performed a spatial delayed-response task for liquid or food reward. The monkeys kept their right hands relaxed on an unmovable resting key. The monkeys faced a 13-inch computer screen positioned behind a transparent plastic wall in which two small levers were mounted to the left and right of the midline. To start each trial, a colour instruction picture ($13^\circ \times 13^\circ$) appeared for 1 s on a computer screen above the left or right lever (Fig. 1). After a randomly varied delay of 2.5–3.5 s, two identical red squares appeared simultaneously as movement-trigger stimuli at the left and right positions of the instruction. The moderately fluid- or food-deprived animal released a key, touched the lever at the position previously indicated by the instruction and received a liquid or food reward for correct performance. Both squares disappeared upon lever touch.

Liquid rewards (0.15 ml) were dispensed 1.5 or 2.0 s after lever touch by a computer-controlled liquid valve from a spout at the animal's mouth. Liquids were grenadine and apple juice for one monkey, and orange and grape juice for the second monkey. Food rewards were presented 2.0 s after lever touch in a box located to the right of the computer screen following computer-controlled opening of its door (40 mm $\times$ 40 mm frontal opening). Foods were raisins (most preferred), small apple morsels (intermediately preferred) and sugar-honey cereal (least preferred). Each reward was indicated at trial onset by a specific instruction picture. To assess the influence of visual features on neuronal responses, we used five different pairs of instruction pictures in liquid trials (four of which are shown in Fig. 3) and one set of three pictures in food trials (Fig. 4). Only two instruction pictures, with their associated two liquid or two food rewards, were presented in a given block of trials.

Reward preferences were assessed in separate blocks of choice trials before or after recording from each neuron. Two different instructions for two rewards were shown simultaneously at randomly alternating left and right target

positions, allowing the animal to touch the lever of its choice following the trigger stimulus. All rewards were used in combinations in which animals showed reliable and persistent preferences. Thus, each pair of instruction stimuli contained one picture associated with a preferred reward and one with a non-preferred reward. Rewards and target positions alternated randomly, with a maximum of three consecutive identical trials. Trials lasted 12–14 s; intertrial intervals were 4–6 s.

Electrophysiological recording. The activity of single neurons in the left orbitofrontal cortex was recorded extracellularly with movable microelectrodes for 20–60 min in the two monkeys, using standard electrophysiological techniques. During neuronal recordings, arm-muscle activity, mouth-muscle activity and eye movements were monitored through chronically implanted electrodes, and licking movements were monitored with an infrared-light barrier interrupted by the animal's tongue. Following analog-to-digital conversion, muscle activity and eye movements were exhibited in single-trial or averaged mode in reference to individual task events. Task-related neuronal changes were assessed with the sliding-window procedure based on the one-tailed Wilcoxon test ($P < 0.01$)^{16,18}. Task-related activations were considered only from neurons tested in at least ten trials in a given situation and showing statistically significant activity increases in relation to at least one task event, compared with the spontaneous activity before the first task event. Task-related changes in individual neurons in response to first, different rewards, second, left and right targets, and third, corresponding pictures of different instruction sets were compared using the two-tailed Mann–Whitney U -test ($P < 0.05$), on impulse counts in single trials. Reaction times (from trigger stimulus onset to key release) in response to different rewards were compared using the two-tailed Mann–Whitney U -test ($P < 0.002$). Recording sites were marked with small electrolytic lesions and reconstructed from 40- μ m-thick cresyl-violet-stained coronal sections of paraformaldehyde-perfused brains. Experimental protocols conformed to the Swiss Animal Protection Law and were supervised by the Fribourg Cantonal Veterinary Office.

Received 2 December 1998; accepted 10 February 1999.

1. Damasio, A. R. *Descartes Error* (Putnam, New York, 1994).
2. Rolls, E. T. The orbitofrontal cortex. *Phil. Trans. R. Soc. Lond. B* **351**, 1433–1444 (1996).
3. Bechara, A., Damasio, H., Tranel, D. & Anderson, S. W. Dissociation of working memory from decision making within the human prefrontal cortex. *J. Neurosci.* **18**, 428–437 (1998).
4. Iversen, S. D. & Mishkin, M. Preservative interference in monkeys following selective lesions of the inferior prefrontal convexity. *Exp. Brain Res.* **11**, 376–386 (1970).
5. Dias, R., Robbins, T. W. & Roberts, A. C. Dissociation in prefrontal cortex of affective and attentional shifts. *Nature* **380**, 69–72 (1996).
6. Baylis, L. L. & Gaffan, D. Amygdectomy and ventromedial prefrontal ablation produce similar deficits in food choice and in simple object discrimination learning for an unseen reward. *Exp. Brain Res.* **86**, 617–622 (1991).
7. Dickinson, A. & Balleine, B. Motivational control of goal-directed action. *Anim. Learn. Behav.* **22**, 1–18 (1994).
8. Jacobsen, C. F. & Nissen, H. W. Studies of cerebral function in primates: IV. The effects of frontal lobe lesions on the delayed alternation habit in monkeys. *J. Comp. Physiol. Psychol.* **23**, 101–112 (1937).
9. Thorpe, S. J., Rolls, E. T. & Maddison, S. The orbitofrontal cortex: neuronal activity in the behaving monkey. *Exp. Brain Res.* **49**, 93–115 (1983).
10. Schoenbaum, G., Chiba, A. A. & Gallagher, M. Orbitofrontal cortex and basolateral amygdala encode expected outcome during learning. *Nature Neurosci.* **1**, 155–159 (1998).
11. Critchley, H. G. & Rolls, E. T. Hunger and satiety modify the responses of olfactory and visual neurons in the primate orbitofrontal cortex. *J. Neurophysiol.* **75**, 1673–1686 (1996).
12. Critchley, H. G. & Rolls, E. T. Olfactory neuronal responses in the primate orbitofrontal cortex: analysis in an olfactory discrimination task. *J. Neurophysiol.* **75**, 1659–1672 (1996).
13. Schoenbaum, G. & Eichenbaum, H. Information coding in the rodent prefrontal cortex. I. Single-neuron activity in orbitofrontal cortex compared with that in pyriform cortex. *J. Neurophysiol.* **74**, 733–750 (1995).
14. Rolls, E. T., Critchley, H. D., Mason, R. & Wakeman, E. A. Orbitofrontal cortex neurons: role in olfactory and visual association learning. *J. Neurophysiol.* **75**, 1970–1981 (1996).
15. Tinklepaugh, O. L. An experimental study of representation factors in monkeys. *J. Comp. Psychol.* **8**, 197–236 (1928).
16. Hollerman, J. R., Tremblay, L. & Schultz, W. Influence of reward expectation on behavior-related neuronal activity in primate striatum. *J. Neurophysiol.* **80**, 947–963 (1998).
17. Hikosaka, O., Sakamoto, M. & Usui, S. Functional properties of monkey caudate neurons. III. Activities related to expectation of target and reward. *J. Neurophysiol.* **61**, 814–832 (1989).
18. Schultz, W., Apicella, P., Scarnati, E. & Ljungberg, T. Neuronal activity in monkey ventral striatum related to the expectation of reward. *J. Neurosci.* **12**, 4595–4610 (1992).
19. Fuster, J. M. & Alexander, G. E. Neuron activity related to short-term memory. *Science* **173**, 652–654 (1971).
20. Kubota, K. & Niki, H. Prefrontal cortical unit activity and delayed alternation performance in monkeys. *J. Neurophysiol.* **34**, 337–347 (1971).
21. Funahashi, S., Chafee, M. V. & Goldman-Rakic, P. S. Prefrontal neuronal activity in rhesus monkeys performing a delayed anti-saccade task. *Nature* **365**, 753–756 (1993).
22. Wilson, F. A. W., O'Scalaidhe, S. P. & Goldman-Rakic, P. S. Dissociation of object and spatial processing domains in primate prefrontal cortex. *Science* **260**, 1955–1958 (1993).
23. Rao, S. C., Rainer, G. & Miller, E. K. Integration of what and where in the primate prefrontal cortex. *Science* **276**, 821–824 (1997).
24. Miller, E. K., Erickson, C. A. & Desimone, R. Neural mechanisms of visual working memory in prefrontal cortex of the monkey. *J. Neurosci.* **16**, 5154–5167 (1996).

25. Petrides, M. & Pandya, D. N. Association fiber pathways to the frontal cortex from the superior temporal region in the rhesus monkey. *J. Comp. Neurol.* **273**, 52–66 (1988).
26. Pandya, D. N. & Yeterian, E. H. Comparison of prefrontal architectures and connections. *Phil. Trans. R. Soc. Lond. B* **351**, 1423–1432 (1996).
27. Barbas, H. Anatomic organization of basoventral and mediadorsal visual recipient prefrontal regions in the rhesus monkey. *J. Comp. Neurol.* **276**, 313–342 (1988).
28. Carmichael, S. T. & Price, J. L. Limbic connections of the orbital and medial prefrontal cortex in macaque monkeys. *J. Comp. Neurol.* **363**, 615–641 (1995).
29. Rosenkilde, C. E., Bauer, R. H. & Fuster, J. M. Single cell activity in ventral prefrontal cortex of behaving monkeys. *Brain Res.* **209**, 375–394 (1981).
30. Watanabe, M. Reward expectancy in primate prefrontal neurons. *Nature* **382**, 629–632 (1996).

Acknowledgements. We thank B. Aebischer, J. Corpataux, A. Gaillard, A. Pisani, A. Schwarz and F. Tinguely for technical assistance. The study was supported by the Swiss NSF (W.S.) and the Fondation pour la Recherche Scientifique de Quebec (L.T.).

Correspondence and requests for materials should be addressed to W.S. (e-mail: Wolfram.Schultz@unifr.ch).

p63 is a p53 homologue required for limb and epidermal morphogenesis

Alea A. Mills^{*†}, Binhai Zheng[†], Xiao-Jing Wang^{‡§}, Hannes Vogel^{||}, Dennis R. Roop^{‡§} & Allan Bradley^{*†}

^{*} Howard Hughes Medical Institute, and [†] Departments of Molecular and Human Genetics, [‡] Cell Biology and [§] Dermatology, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030, USA

^{||} Department of Pathology, Texas Childrens Hospital, Houston, Texas 77030, USA

The p53 tumour suppressor is a transcription factor that regulates the progression of the cell through its cycle and cell death (apoptosis) in response to environmental stimuli such as DNA damage and hypoxia^{1,2}. Even though p53 modulates these critical cellular processes, mice that lack p53 are developmentally normal³, suggesting that p53-related proteins might compensate for the functions of p53 during embryogenesis. Two p53 homologues, p63 and p73, are known^{4,5} and here we describe the function of p63 *in vivo*. Mice lacking p63 are born alive but have striking developmental defects. Their limbs are absent or truncated, defects that are caused by a failure of the apical ectodermal ridge to differentiate. The skin of p63-deficient mice does not progress past an early developmental stage: it lacks stratification and does not express differentiation markers. Structures dependent upon epidermal–mesenchymal interactions during embryonic development, such as hair follicles, teeth and mammary glands, are absent in p63-deficient mice. Thus, in contrast to p53, p63 is essential for several aspects of ectodermal differentiation during embryogenesis.

Two p53-related genes, human p73 and rat *Ket*, have been described^{4,5}. We cloned a mouse p53 homologue referred to here as p63, using a polymerase chain reaction strategy designed to identify genes that contain homology to the highly conserved DNA-binding domain of p53. Sequence analysis indicates that this gene contains significant homology to p53, human p73 and rat *Ket*. p63 was mapped to mouse chromosome 16 between *D16Mit1* and *D16Mit3* (data not shown), which indicated that p63 was not the mouse homologue of human p73. Complementary DNAs encoding the human homologue of *Ket* have recently been described and designated as p40 (ref. 6), p51 (ref. 7), and p63 (ref. 8). Sequence alignments of rat *Ket*, mouse p63 and human p63 indicate that these genes are true homologues (data not shown).

Expression of p63 was examined using *in situ* and northern-blot hybridization. p63 is expressed as early as embryonic day 9.5 (E9.5) within the oral ectoderm, limb buds and tail bud region (see below). At later stages of gestation, p63 is expressed primarily within the ectoderm; expression is evident within the basal region of the interfollicular epidermis of the skin and within the outer root-

sheath of hair follicles (data not shown). In adult mice, northern-blot analysis indicates that p63 is expressed in skin, tongue, tail and skeletal muscle (data not shown).

To assess the function of p63 during embryogenesis, we mutated the p63 gene using embryonic stem (ES) cell technology. Two different targeting vectors, pTV6H(90) and pTV12E(60), were identified that integrate into different regions of the p63 locus, and these were used to generate different mutant alleles of the p63 gene (Fig. 1a). These vectors were isolated from a library of pre-prepared vectors³¹ which use the gap-repair mechanism in their gene-targeting reactions⁹; an internal fragment was deleted from the region of homology for use as a diagnostic probe to detect gene targeting. The pTV6H(90) vector generates a recombinant allele that is predicted to truncate the transcript within exon 6 and which, if translated, would create a protein with a non-functional DNA-binding domain; this allele is referred to as p63^{Brdm1}. The pTV12E(60) vector generates a recombinant allele, p63^{Brdm2}, that truncates the messenger RNA at exon 10: this transcript (if translated) would yield a protein containing an intact DNA-binding domain, but not the highly conserved 3' region of the p63 protein. Targeted ES cell clones were identified and chimaeras were generated by blastocyst microinjection. Multiple independent ES clones representing each class of recombinant allele transmitted their mutant p63 alleles into the germ line.

F₁(129/C57B6) heterozygous mutant mice were intercrossed, ~25% of the pups from these matings were born with a shiny, transparent skin (Fig. 2a). These pups had abnormal facies, truncations of the forelimbs, and were without hindlimbs. Determination of the genotype of mice with this phenotype showed that they were exclusively homozygous mutants (Fig. 1b). Northern-blot analysis indicated that p63-homozygous mutant animals do not express the p63 transcript (Fig. 1c), so these are likely to be null alleles. Independent clones representing the two different mutant alleles p63^{Brdm1} and p63^{Brdm2} produced an identical phenotype (Fig. 2a) and therefore were not distinguished in subsequent experiments.

Although p63-deficient animals were viable at birth, they died several hours later. Organ dissection from the abdominal and thoracic cavities of p63-deficient newborn mice showed no deviation from the usual *situs* arrangement or any gross malformation, and histological sections of the viscera and brain revealed no microscopic abnormality. Water-loss assays, an *in vivo* measure of the functional permeability of the skin, revealed that p63-deficient animals lose approximately thirty times more water than their wild-type littermates (data not shown). It is likely that these p63-deficient animals die from dehydration.

p63-deficient newborns display striking limb defects. The forelimbs were truncated and hindlimbs were completely absent in all of the p63-homozygous mutant animals that were analysed visually at birth (*n* = 77). Analysis of skeletons stained for bone and cartilage indicates that the forelimbs of p63-homozygous mutant animals are truncated and that the distal skeletal elements are absent (Fig. 2b and c). Phalanges and carpals were absent in all of the p63-homozygous mutant forelimb skeletal preparations analysed, whereas more proximal forelimb structures were slightly heterogeneous in the extent of the truncation. For example, the radius was not present in any of the mutant limbs analysed, but the ulna was present in a subset (37.5%) of the limbs. Although the humerus was present in each of the mutant limbs, it was usually truncated, deformed, and smaller in girth than those of wild-type and heterozygous siblings (Fig. 2c). The femur and all distal skeletal elements were absent in all of the p63-homozygous mutant limbs examined (Fig. 2b, d). The hip girdle was present but lacked the ossification centre found in the wild-type sibling. Teeth are also absent in p63-deficient newborn mice (Fig. 2e).

During embryogenesis, the apical ectodermal ridge (AER), a structure required for limb outgrowth along the proximal–distal axis, can be seen in the scanning electron micrograph at the junction

of the dorsal and ventral surfaces of the distal tip of the limbs of E11.5 wild-type embryos (Fig. 3). In contrast, the limb buds of *p63*-deficient embryos are distinctly smaller and misshapen, and there is no morphologically distinct AER. Histological analysis of the limb buds of E10.5 embryos confirmed that the AER is absent in *p63*-homozygous mutant limbs (Fig. 3). Analysis of the pattern of expression of *p63* by *in situ* hybridization of whole mounts and sections indicates that *p63* is normally expressed within the ectoderm of the branchial arches, tail and limbs (Fig. 4a).

To investigate the molecular basis underlying these morphological defects, we used whole-mount *in situ* hybridization to examine E10.5 embryos for the expression of several genes that are important in limb-bud outgrowth (Fig. 4b). *Lmx-1*, a marker of the dorsal limb mesenchyme¹⁰, is expressed abnormally within the limbs of *p63*-

homozygous mutant embryos. The dorsal–ventral border of *Lmx-1* expression that is seen in the E10.5 wild-type embryo is perturbed in the *p63*-homozygous mutant sibling, and *Lmx-1* expression extends into the ventral mesenchyme¹⁰. *Fgf-8*, a marker of the AER¹¹, is not detectable in the limbs of *p63*-homozygous mutant embryos at E10.5. To investigate whether the *p63*-deficient limb phenotype is caused by a defect in ectodermal–mesenchymal signalling, we analysed embryos for the expression of the homeodomain-containing transcription factors *Msx-1* and *Msx-2*. Expression of *Msx-1* in the mesenchyme depends on an ectodermal signal, whereas expression of *Msx-2* is independent of signalling from the ectoderm¹². *p63*-homozygous mutant limb buds express *Msx-2* in a pattern similar to that of wild-type embryos. In contrast, *Msx-1* is not expressed in the mesenchyme of *p63*-deficient limbs, indicating that

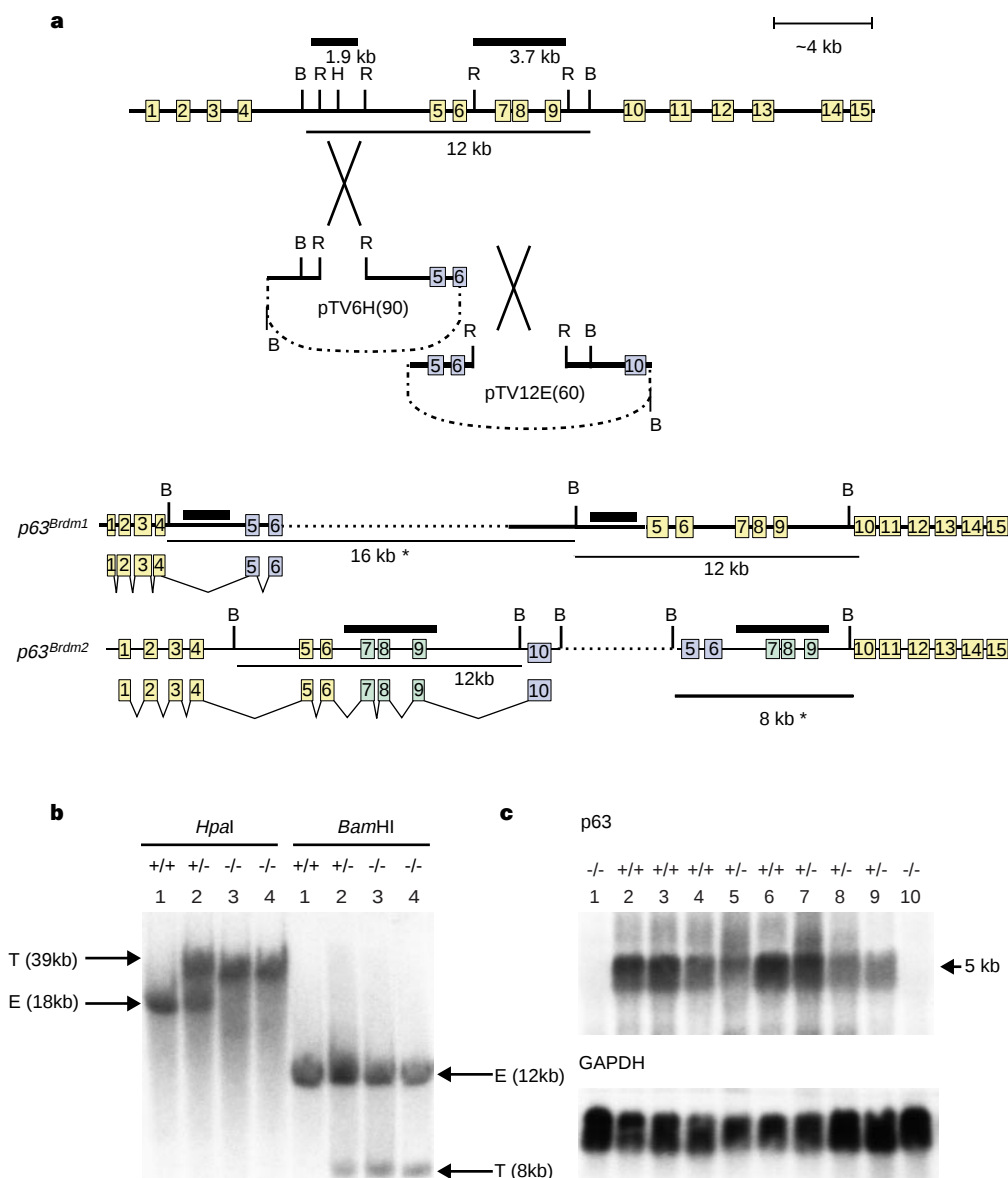


Figure 1 Targeted disruption of the *p63* locus. **a**, Diagram of the *p63* locus is shown (top) with exons depicted as yellow boxes. Two different insertion vectors containing the *p63* exons designated (purple boxes) were used to disrupt *p63* (middle). Targeted loci *p63^{Brdm1}* and *p63^{Brdm2}* are shown (bottom), with repaired exons represented as green boxes. Targeting was assessed by Southern blotting using 1.9- and 3.7-kb *EcoRI* probes (bold lines) to detect 16- and 8-kb diagnostic fragments within *p63^{Brdm1}* and *p63^{Brdm2}* alleles, respectively (asterisk). Restriction

enzyme cleavage sites: B, *Bam*HI; H, *Hpa*I; R, *Eco*RI. **b**, Southern blot hybridization of DNA obtained from the progeny of *p63^{Brdm2}/+* intercrosses was used to detect endogenous (E) and targeted (T) *p63* alleles. **c**, Northern blot hybridization of total RNA isolated from a litter of newborn mice from a heterozygous intercross. A *p63* transcript is not expressed in *p63^{Brdm2}/p63^{Brdm2}* siblings. A *GAPDH* (glyceraldehyde-3-phosphate dehydrogenase) probe was used as a control.

ectodermal–mesenchymal signalling is perturbed in *p63*-homozygous mutant limbs. Appropriate interactions between the ectoderm and the mesenchyme are essential for AER formation, and perturbation of these interactions underlies the limb defect observed in the chick mutant *limbless*¹³. The phenotype observed in *p63*-deficient limbs is similar to the *limbless* mutant in several respects. Both of these mutants express *Msx-2* in the limb-bud mesenchyme, but do not appear to express *Msx-1* and *Fgf-8*. Furthermore, no morphologically distinct AER is detected in either of these mutants. As is the case for the *limbless* mutant, the limb truncations of *p63*-null mice are likely to be due to a defect in ectodermal–mesenchymal interactions, resulting in a lack of AER formation at the distal tip of the limb bud at the dorsal–ventral boundary of the ectoderm. Indeed, the limb truncations observed in *p63*-null animals are

reminiscent of the limbs that are formed after surgical removal of the AER¹⁴.

Other structures that require ectodermal–mesenchymal interaction during morphogenesis were examined in *p63*-deficient mice, using expression of *Lef-1* as a marker. *Lef-1* is normally expressed within the ectoderm of hair placodes, the precursors of hair follicles, and within mammary buds before morphogenesis, providing a cue to the underlying mesenchyme to initiate development of the appropriate structure¹⁵. In *p63*-deficient mice, *Lef-1* is not expressed within the ectoderm overlying the site at which whisker follicles (Fig. 4c) and mammary buds normally form (data not shown), suggesting that ectodermal–mesenchymal signalling is perturbed.

At birth, *p63*-deficient mice have striking and visible skin defects. Histological analysis of neonatal *p63*-deficient skin revealed the

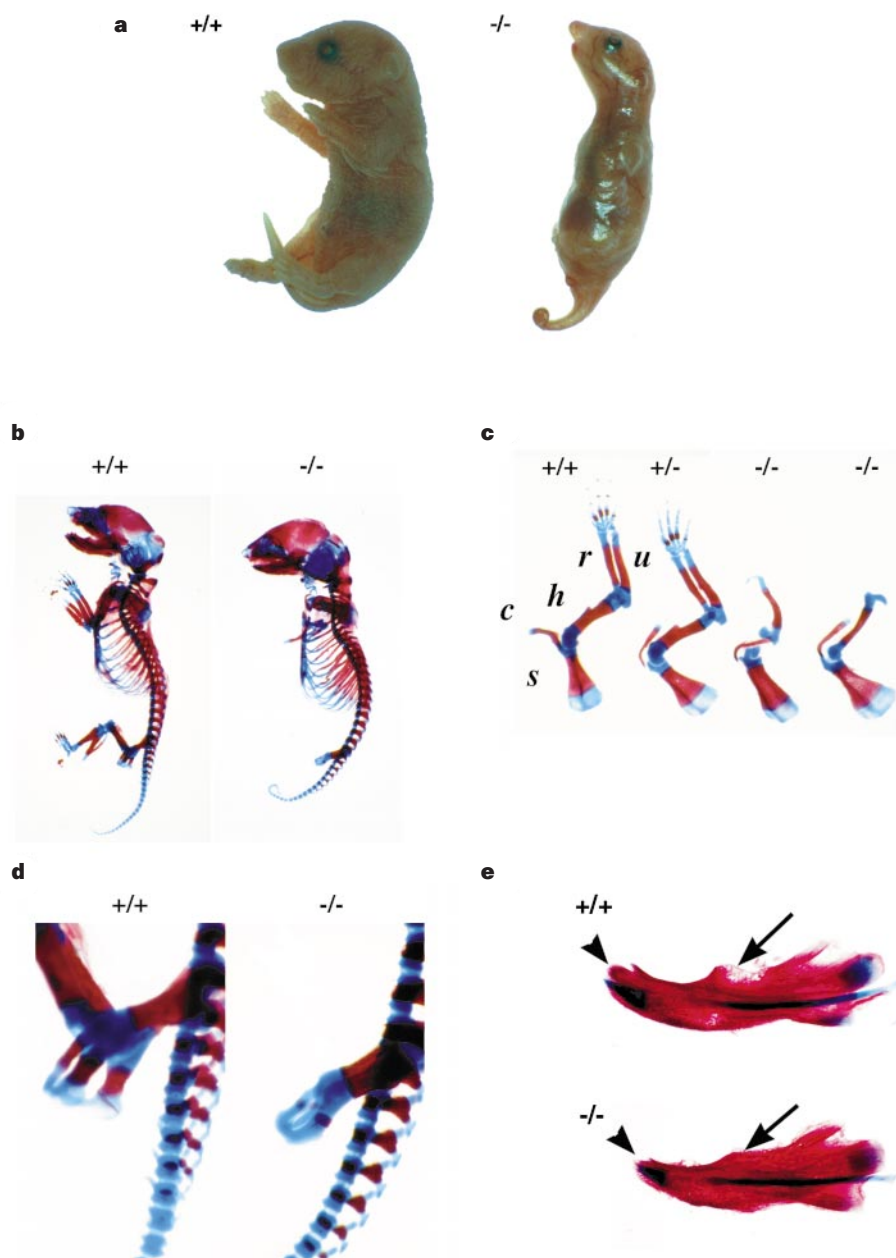


Figure 2 The phenotype of *p63*-deficient newborn mice. **a.**, Matings between *p63*-heterozygous mice produce wild-type and heterozygous offspring that are overtly normal and *p63*-deficient mice that have severe limb and skin defects. **b.**, Skeletons of wild-type, *p63*-heterozygous and *p63*-homozygous mutant animals were stained for cartilage (blue) and bone (red). **c.**, Forelimbs of *p63*-homozygous

mutant mice are truncated; the phalanges, radius and ulna are absent and the humerus is deformed. Abbreviations: c, clavicle; h, humerus; r, radius; s, scapula; u, ulna. **d.**, Hindlimbs are absent in *p63*-deficient animals. **e.**, Incisors (arrowheads) and a prominent alveolar ridge (arrows) are not present in *p63*-deficient mice.

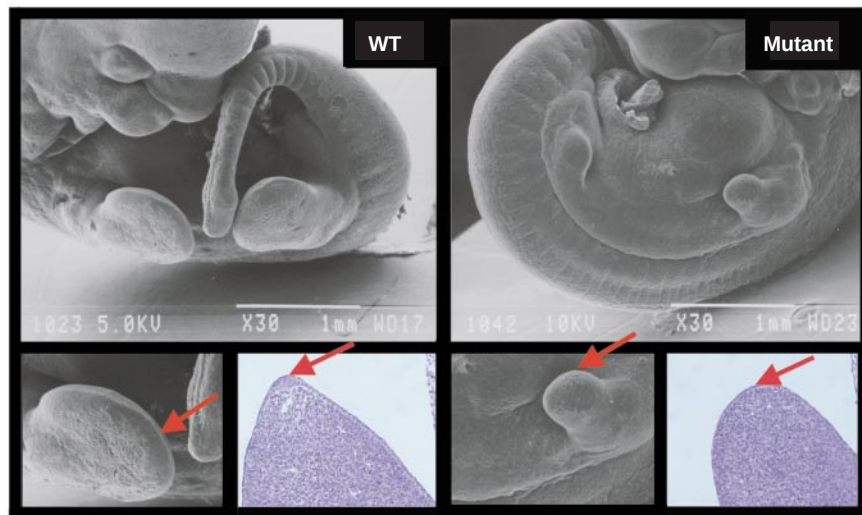


Figure 3 Analysis of *p63*-deficient limbs during embryogenesis. Scanning electron microscopy and histological analysis of wild-type (WT, left) and *p63*-deficient (mutant, right) embryos indicates that the AER is absent in *p63*-homozygous mutant embryos (arrows).

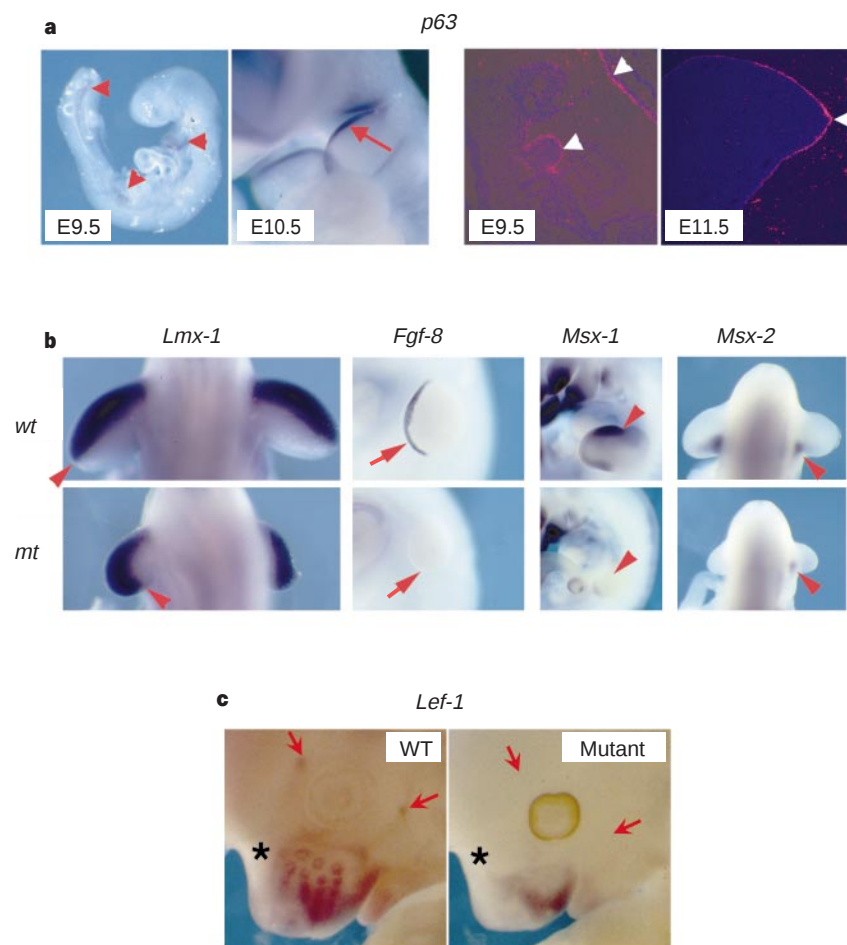


Figure 4 Analysis of expression in wild-type and *p63*-deficient limbs. **a**, Expression of *p63* using whole-mount (left) and section (right) *in situ* hybridization. At E9.5, expression is detectable in the branchial arch and the forelimb bud, and within the tailbud region (arrowheads). At E10.5, expression of *p63* is evident in the oral epithelium (arrowheads). *p63* is expressed within the AER (arrowhead) with the dorsal and ventral ectoderm of developing limbs at E11.5. **b**, E10.5 embryos were analysed for expression of *Lmx-1*, *Fgf-8*, *Msx-1* and *Msx-2*. *p63*-deficient limbs do not have an appropriate dorsal-ventral boundary of *Lmx-1* expression (arrowheads). *Fgf-8* is not detectable in the limbs of *p63*-deficient embryos at E10.5 (red arrows). Expression of *Msx-1* is absent, whereas expression of *Msx-2* is normal in the limbs of *p63*-deficient animals (arrowheads). **c**, Analysis of E12.5 embryos for *Lef-1* expression. Note that *Lef-1* is expressed in the wild-type embryo but not in the mutant in the whisker pad (asterisks), supraorbital and suborbital follicles (arrows).

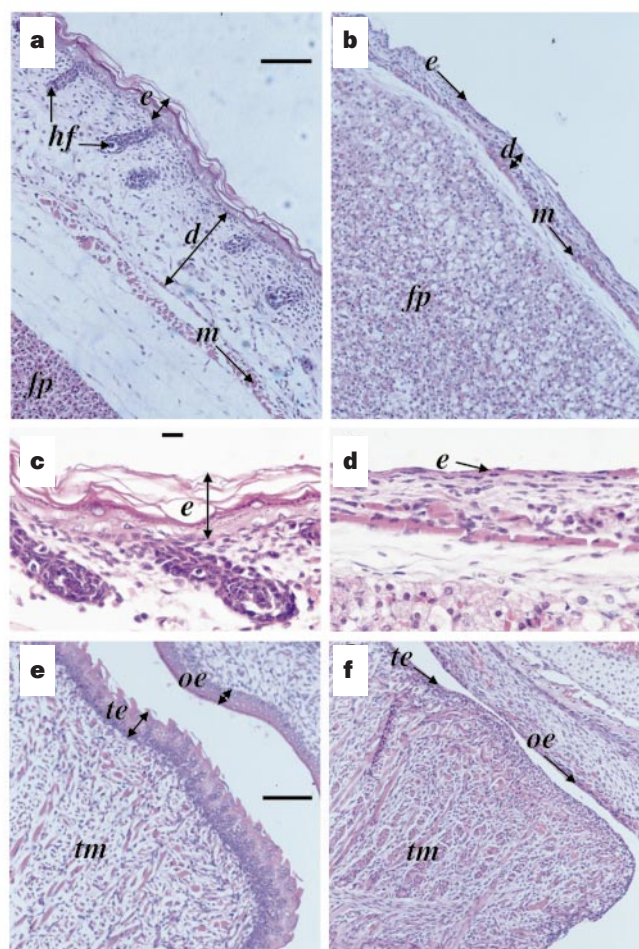


Figure 5 *p63* is required for development of the epidermis. Histological analysis of dorsal back skin (**a–d**) and the oral cavity (**e, f**) of wild-type (**a, c, e**) and *p63*-deficient (**b, d, f**) newborns indicates that epithelial differentiation is perturbed in *p63*-deficient mice (**a** and **b** are shown at higher magnification in **c** and **d**). The thickness of the dermis and epidermis is reduced and hair follicles are absent in the skin of *p63*-null mice (**b, d**) as compared to the skin of wild-type siblings (**a, c**). Tongue and oral epithelia are abnormal in *p63*-homozygous mutant mice (**f**) and lack stratification and the presence of differentiated structures seen in the wild-type animal (**e**). **d**, Dermis; **e**, epidermis; **fp**, fat pad; **hf**, hair follicle; **m**, muscle; **oe**, oral epithelium; **te**, tongue epithelium; **tm**, tongue muscle.

absence of the normal epidermal structure and complete lack of hair follicles (Fig. 5a–c). The surface of *p63*-deficient skin is covered by a single layer of flattened cells, without the spinosum, granulosum and stratum corneum. Likewise, the epithelia of the tongue and oral cavity of *p63*-deficient mice consist of only a single cell layer, without the differentiated structure that is present in wild-type oral epithelia (Fig. 5e and f).

To determine the stage at which epidermal development ceases in the skin of *p63*-homozygous mutants, several differentiation markers were examined using immunofluorescence (Fig. 6a). The keratin-14 (*K14*) gene is normally expressed within proliferating keratinocytes and its protein product persists into the upper layers of the epidermis¹⁶. *K14* was expressed at a very low level in the single layer of *p63*-deficient epidermis. Keratin-1 (*K1*), an early marker of epidermal differentiation¹⁶, is detected in the epidermis of wild-type and *p63*-heterozygous mutant mice throughout the suprabasal layers of the epidermis, and in some postmitotic basal cells, but is not detectable in *p63*-homozygous mutant skin. Furthermore,

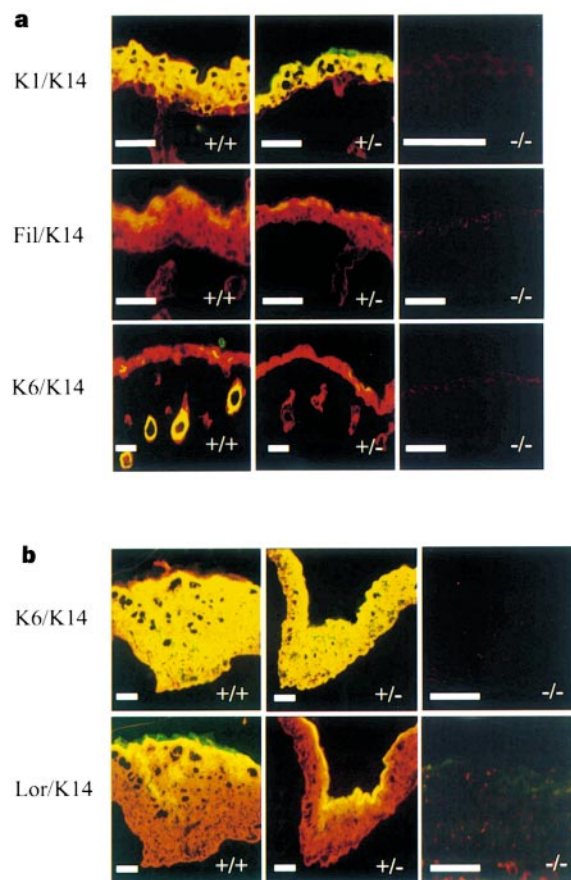


Figure 6 Expression of differentiation markers in *p63*-deficient epidermis. **a, b**, Double-staining immunofluorescence of newborn skin (**a**) and oral cavity (**b**) of wild-type (+/+), *p63*-heterozygous mutant (+/-), and *p63*-homozygous mutant (-/-) mice. **a**, Skin sections were stained for expression of keratin-14 (*K14*, red). Keratin-1 (*K1*), filaggrin (*Fil*) and keratin-6 (*K6*) are stained green; where expression overlaps with that of *K14*, expression appears yellow. *K14* is weakly expressed in *p63*-homozygous epidermis, and *K1*, *Fil* and *K6* are not detectable. Although *p63*-heterozygous mice express *K14*, *K1* and filaggrin, the overall thickness of the epidermis is less than that of wild-type mice. *K6* is not expressed in the *p63*-heterozygous animals. **b**, Sections of the oral cavity were stained for expression of *K14* in combination with either *K6* or loricrin (*Lor*), as in **a**. *p63*-deficient mice fail to express *K6* and loricrin, and *K14* expression is only weakly detected. Note that the thickness of the oral epithelium in *p63*-heterozygous mice is approximately half that of the wildtype sibling. Scale bar, 32 μ m.

several other epidermal differentiation markers are also absent in *p63*-deficient skin, including the early-differentiation marker keratin-10 (ref. 16) (data not shown) and the late-stage terminal differentiation markers filaggrin¹⁷ (Fig. 6a) and loricrin¹⁸ (data not shown). The pattern of expression of differentiation markers was also examined in oral epithelia (Fig. 6b). *K14* and *K6* are detected throughout all layers of normal oral epithelia, whereas expression of keratin-13 (*K13*) and loricrin are restricted to the differentiated layers. None of these markers are detectable in the oral epithelia of homozygous mutants (Fig. 6b and data not shown).

In normal epidermis, *K14* expression is first detected at E9.5 (ref. 19), hair follicles begin to develop around E14.5 (ref. 20), and stratification is evident at E15.5 (ref. 21). As the epidermis of *p63*-homozygous mutant animals exhibits only faint and patchy staining of *K14*, does not form hair follicles, and lacks stratified structure, epidermal development apparently ceased at approximately E9.5.

Targeted disruption of *p63* has revealed that this gene plays an essential role in both morphogenesis of the epidermis and limb

development. Several mutations in the chick^{22,23} and mouse²⁴ cause severe limb truncations. However, the striking defects in epidermal morphogenesis seen in p63-deficient mice are not found in any of these mutants. The downstream targets of p63 probably involve direct or indirect activation of factors involved in ectodermal-mesenchymal communication, which are required for morphogenesis of the hair follicle, tooth, mammary bud and AER. p63 is the first p53 homologue with a clear role in development. Whether p63 functions as a tumour suppressor is under investigation. □

Methods

Cloning and mapping of the p63 gene. Primers F (5'-CCCCATCCAGATC AAGGTGATG-3' and G (5'-TTGAAGTCGCGGCTCAGCTCGT-3') were used to clone the mouse p63 gene from the 3' hprt gene-targeting vector library³¹. This library was constructed with mouse 129S5 genomic DNA using a vector backbone that contains the puromycin-resistance gene, the K14/Agouti gene and the non-functional 3' portion of the hprt cassette²⁵. Two overlapping bacteriophage lambda clones encompassing exons 5–10 of the mouse p63 gene were isolated. Clone 6H(90) contains an 8.3-kb genomic insert encoding exon 5 and a portion of exon 6; clone 12E(60) contains a 10-kb genomic insert which encodes exons 5–10. Screening of a C57BL/6J mouse genomic BAC library²⁶ with an exon-5 probe led to the isolation of two overlapping BACs which encompass the p63 locus.

The p63 gene was mapped using the 3.7-kb EcoRI probe from clone 12E(60), which detects an SstI restriction fragment length polymorphism between C57BL/6J and *Mus spretus*. This probe was hybridized to filters of SstI-digested DNA from the BSS backcross panel (Jackson Laboratories).

Construction of targeting vectors and targeted disruption of the p63 locus. The pTV6H(90) and pTV12E(60) targeting vectors were generated by deleting a 1.9- and a 3.7-kb EcoRI fragment from clone 6H(90) and 12E(60), respectively. 20 µg of EcoRI-linearized constructs was electroporated into AB2.2 ES cells, cells were plated on feeder layers of gamma-irradiated SNL76/7 fibroblasts, integration of the targeting construct was positively selected for by growth in puromycin, and clones were picked after 8–10 days of selection and transferred to 96-well feeder plates. Targeting of the p63 locus was analysed by Southern blotting. Clones transfected with the pTV12E(60) targeting vector were screened using a BglII digest and the 3.7-kb EcoRI fragment as a probe; clones transfected with pTV6H(90) targeting construct were screened with a BamHI digest using the 1.9-kb EcoRI fragment as a probe.

Transmission of p63^{Brdm1} and p63^{Brdm2} mutant alleles. Targeted ES-cell clones were injected into E3.5 blastocysts of the C57BL/6J-Tyr^{c-Brd} mouse strain²⁷ using standard procedures²⁸. Injected blastocysts were transferred to the uteri of pseudopregnant female B6/CBA F₁ recipients. The resulting chimaeras were mated with C57BL/6J-Tyr^{c-Brd} females to establish the p63 mutant alleles in the germ line. Heterozygous progeny were identified visually from their coat colour and confirmed by Southern analysis. BamHI digests of tail DNA were used to differentiate wild-type and p63^{Brdm1} alleles; HpaI digests were used to differentiate between heterozygous and homozygous animals.

Skeletal preparations. Newborn pups or embryos obtained from timed matings were killed and eviscerated, and the skin was removed. Skeletons were fixed overnight at room temperature in 100% ethanol and stained for 48 h in 3 mg ml⁻¹ Alcian blue in 20% glacial acetic acid, 80% ethanol. After a brief rinse in 95% ethanol, the soft tissue was removed by incubating for 24 h in 2% potassium hydroxide, and bone was stained for 24 h in 75 µg ml⁻¹ Alizarin red in 1% potassium hydroxide. Skeletons were cleared in 1% KOH, 20% glycerol for 2–3 weeks, washed for 18 h in 10% glycerol, 20% ethanol, and placed in 50% glycerol, 50% ethanol for photography and storage.

Scanning electron microscopy. Embryos were fixed overnight at 4 °C in 10% neutral buffered formalin and dehydrated in a series of 50, 70, 90, 96 and 100% ethanol. Embryos were critical-point-dried, coated with a mixture of 60% gold and 40% palladium in an argon chamber, and visualized using a JSM-6100 scanning electron microscope.

In situ hybridization. Whole-mount and section *in situ* hybridization was performed as described²⁹ on embryos obtained from matings timed from the observation of a vaginal plug, estimated to be at 0.5 days post coitum (E0.5).

Histological analysis. Tissues were fixed in 10% formalin, neutral buffered

saline dehydrated, embedded in paraffin, and sectioned at 5 µm thickness. Paraffin was removed and the tissues were rehydrated, stained in haematoxylin, counterstained in eosin, and dehydrated in an ethanol series ending in xylene.

Immunofluorescence. Tissues were collected, embedded in OCT and frozen on dry ice. Tissue sections were stained using antibodies specific for K1, K6, K10, K13, K14, filaggrin and loricrin. Double-staining immunofluorescence was used to detect expression of a given marker in conjunction with K14 as described³⁰.

Received 11 January; accepted 3 March 1999.

- Livingston, L. R. *et al.* Altered cell cycle arrest and gene amplification potential accompany loss of wild-type p53. *Cell* **70**, 923–935 (1992).
- Lowe, S. W., Schmitt, E. M., Smith, S. W., Osborne, B. A. & Jacks, T. p53 is required for radiation-induced apoptosis in mouse thymocytes. *Nature* **362**, 847–849 (1993).
- Donehower, L. A. *et al.* Mice deficient for p53 are developmentally normal but susceptible to spontaneous tumours. *Nature* **356**, 215–221 (1992).
- Kaghad, M. *et al.* Monoallelically expressed gene related to p53 at 1p36, a region frequently deleted in neuroblastoma and other human cancers. *Cell* **90**, 809–819 (1997).
- Schmale, H. & Bamberger, C. A novel protein with strong homology to the tumor suppressor p53. *Oncogene* **15**, 1363–1367 (1997).
- Trink, B. *et al.* A new human p53 homologue. *Nature Med.* **4**, 747–748 (1998).
- Osada, M. *et al.* Cloning and functional analysis of human p51, which structurally and functionally resembles p53. *Nature med.* **4**, 839–843 (1998).
- Yang, A. *et al.* p63, a p53 homolog at 3q27–29, encodes multiple products with transactivating, death-inducing, and dominant-negative activities. *Mol. Cell* **2**, 305–316 (1998).
- Valancius, V. & Smithies, O. Double-strand gap repair in a mammalian gene targeting reaction. *Mol. Cell. Biol.* **11**, 4389–4397 (1991).
- Cygan, J. A., Johnson, R. L. & McMahon, A. P. Novel regulatory interactions revealed by studies of murine limb pattern in Wnt-7a and En-1 mutants. *Development* **124**, 5021–5032 (1997).
- Voge, A., Rodriguez, C. & Izpisua-Belmonte, J. C. Involvement of FGF-8 in initiation, outgrowth and patterning of the vertebrate limb. *Development* **122**, 1737–1750 (1996).
- Wang, Y. & Sassoon, D. Ectoderm-mesenchyme and mesenchyme-mesenchyme interactions regulated Msx-1 expression and cellular differentiation in the murine limb bud. *Dev. Biol.* **168**, 374–382 (1995).
- Grieshammer, U., Minowada, G., Pisen, J. M., Abbot, U. K. & Martin, G. R. The chick limbless mutation causes abnormalities in limb bud dorsal-ventral patterning: implications for the mechanism of apical ridge formation. *Development* **122**, 3851–3861 (1996).
- Summerbell, D., Lewis, J. H. & Wolpert, L. Positional information in chick limb morphogenesis. *Nature* **244**, 492–496 (1973).
- Zhou, P., Byrne, C., Jacobs, J. & Fuchs, E. Lymphoid enhancer factor 1 directs hair follicle patterning and epithelial cell fate. *Genes Dev.* **9**, 700–713 (1995).
- Roop, D. R., Krieg, T. M., Mehrel, T., Cheng, C. K. & Yuspa, S. H. Transcriptional control of high molecular weight keratin gene expression in multistage mouse skin carcinogenesis. *Cancer Res.* **48**, 3245–3252 (1988).
- Rothnagel, J. A., Mehrel, T., Idler, W. W., Roop, D. R. & Steinert, P. M. The gene for mouse epidermal flaggrin precursor. Its partial characterization, expression, and sequence of a repeating flaggrin unit. *J. Biol. Chem.* **262**, 15643–15648 (1987).
- Mehrel, T. *et al.* Identification of a major keratinocyte cell envelope protein, loricrin. *Cell* **71**, 1103–1112 (1990).
- Byrne, C., Tainsky, M. & Fuchs, E. Programming gene expression in developing epidermis. *Development* **120**, 2369–2383 (1994).
- Hardy, M. H. & Vielkind, U. Changing patterns of cell adhesion molecules during mouse pelage hair follicle development. 1. Follicle morphogenesis in wild-type mice. *Acta Anal.* **157**, 169–182 (1996).
- Bickenbach, J. R., Greer, J. M., bundman, D. S., Rothnagel, J. A. & Roop, D. R. Loricrin expression is coordinated with other epidermal proteins and the appearance of lipid lamellar granules in development. *J. Invest. Dermatol.* **104**, 405–410 (1995).
- Hinchliffe, J. R. & Ede, D. A. Cell death and the development of limb form and skeletal pattern in normal and wingless (ws) chick embryos. *J. Embryol. Exp. Morphol.* **30**, 753–772 (1973).
- Prahlad, K. V., Skala, G., Jones, D. G. & Briles, W. E. Limbless: a new genetic mutant in the chick. *J. Exp. Zool.* **209**, 427–434 (1979).
- McNeish, J. D., Scott, W. J. Jr & Potter, S. S. Legless, a novel mutation found in PHT1-1 transgenic mice. *Science* **241**, 837–839 (1988).
- Ramirez-Solis, R., Liu, P. & Bradley, A. Chromosome engineering in mice. *Nature* **378**, 720–724 (1995).
- Osoegawa, K. *et al.* An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* **51**, 1–8 (1998).
- Liu, P., Zhang, H., McLellan, A., Vogel, H. & Bradley, A. Embryonic lethality and tumorigenesis caused by segmental aneuploidy on mouse chromosome 11. *Genetics* **150**, 155–168 (1998).
- Bradley, A. in *Teratocarcinomas and Embryonic Stem Cells—a Practical Approach* (ed. Robertson, E. J.) 113–151 (IRI, Oxford, 1987).
- Wilkinson, D. G. *In situ* hybridization. A practical approach. (ed. Wilkinson, D. G.) (Oxford University Press, Oxford, 1992).
- Wang, X. J., Greenhalgh, D. A., Lu, X. R., Bickenbach, J. R. & Roop, D. R. TGF alpha and v-fos cooperation in transgenic mouse epidermis induces aberrant keratinocyte differentiation and stable, autonomous papillomas. *Oncogene* **10**, 279–289 (1995).
- Zheng, B., Mills, A. M. & Bradley, A. A system for rapid generation of coat color tagged knockouts and defined chromosomal rearrangements in mice. *Nucleic Acids Res.* (in the press).

Acknowledgements. We thank S. Rivera for assistance with gene targeting, Y.-C. Cheah for blastocyst injection and for help with mapping the p63 locus, H. Zhang for sequencing, Y. Qi and J. Wesley for assistance in genotyping and maintaining the mouse colony, S. Vaishnav for histology and *in situ* hybridization, R. Johnson for constructs, for advice, and for critically reading the manuscript, and members of A.B.'s laboratory, particularly G. Luo and M. Mills for discussions, P. Biggs for reviewing the text, and S. Perez for help in preparing it. A.B. is an Investigator and A.M. is a Research Associated with the Howard Hughes Medical Institute. A.B. and D.R.R. acknowledge support from the National Cancer Institute.

Correspondence and requests for materials should be addressed to A.B. (e-mail: abradley@bcm.tmc.edu).

p63 is essential for regenerative proliferation in limb, craniofacial and epithelial development

Annie Yang*, Ronen Schweitzer†, Deqin Sun‡, Mourad Kaghad§, Nancy Walker§, Roderick T. Bronson||, Cliff Tabin†, Arlene Sharpe‡¶, Daniel Caput§, Christopher Crum‡ & Frank McKeon*

Departments of * Cell Biology and † Genetics, Harvard Medical School, and ‡ Immunology Research Division, ‡ Department of Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA § Sanofi Biorecherche, Innopole BP 137, 31676 Labège Cedex, France || US Department of Agriculture, Human Nutrition Research Center on Aging, and Department of Pathology, Tufts University School of Veterinary Medicine, Boston, Massachusetts 02111, USA

The *p63* gene, a homologue of the tumour-suppressor *p53* (refs 1–5), is highly expressed in the basal or progenitor layers of many epithelial tissues¹. Here we report that mice homozygous for a disrupted *p63* gene have major defects in their limb, craniofacial and epithelial development. *p63* is expressed in the ectodermal surfaces of the limb buds, branchial arches and epidermal appendages, which are all sites of reciprocal signalling that direct morphogenetic patterning of the underlying mesoderm. The limb truncations are due to a failure to maintain the apical ectodermal ridge, a stratified epithelium, essential for limb development. The embryonic epidermis of *p63*^{−/−} mice undergoes an unusual process of non-regenerative differentiation, culminating in a striking absence of all squamous epithelia and their derivatives, including mammary, lacrimal and salivary glands. Taken together, our results indicate that *p63* is critical for maintaining the progenitor-cell populations that are necessary to sustain epithelial development and morphogenesis.

p63 shows remarkable structural similarity to *p53* and to the related *p73* gene⁶. Unlike *p53*, however, several messenger RNAs are transcribed from two different promoters of the *p63* gene, yielding both transactivating (TA-*p63*) and non-transactivating (Δ N-*p63*) isoforms¹ (Fig. 1a). Embryonic stem cells bearing a *p63* allele lacking exons 6, 7 and 8 were produced by homologous recombination⁷ and used to generate mice heterozygous for this mutation (Fig. 1b, c). These deleted exons comprise most of the DNA-binding domain that is common to all known *p63* isoforms, and are presumably essential for the function of both the TA- and Δ N-*p63* proteins. Immunoblotting of whole-embryo extracts with the 4A4 anti-*p63* antibody¹ showed that *p63*^{−/−} embryos lacked species of relative molecular mass 80K, likely to be the Δ N-63 α isoform, present in both wild-type and heterozygote specimens (Fig. 1d). *p63*^{−/−} mice derived from heterozygous crosses die within a day of birth, owing to desiccation and maternal neglect secondary to severe developmental malformations. A comparison of wild-type and mutant newborn mice and late-stage embryos revealed striking craniofacial abnormalities, limb truncations, and a complete absence of an epidermis and related appendages, including hair follicles, vibrissae, and tooth primordia (Fig. 1f–j).

Defects in limb morphogenesis in *p63* mutants were evident as early as E9.5 and progressed through to E12 (Fig. 2a), a period of limb-bud formation dependent on ectodermal–mesenchymal signalling interactions^{8,9}. During this interval, *p63* expression was evident in the limb buds, branchial arches and the overall ectoderm (Fig. 2b). In particular, the limb buds showed intense staining in the apical ectodermal ridge (AER), a region of specialized, multilayered ectoderm which is essential for growth and patterning of the

underlying mesenchyme^{8–10} (Fig. 2b). The ectodermal surfaces of the branchial arches also showed strong expression of *p63* RNA. This ectoderm is important in directing the morphogenesis of craniofacial skeletal and soft-tissue structures¹¹, both of which are hypoplastic in *p63*-deficient mice. In accord with the *p63* transcript expression patterns, the 4A4 anti-*p63* monoclonal antibody revealed strong nuclear staining of ectodermal cells of E11 embryos, especially those comprising the AER of the limb bud (Fig. 2c) and the maxillary and mandibular branchial ectoderm (results not shown).

The high levels of *p63* expression in the AER suggested a potential role for *p63* in maintaining the structure or function of this specialized ectoderm. We therefore performed serial sectioning of E11 *p63*^{−/−} and control embryos to determine the integrity of the AER. Control sections showed the characteristic, stratified ectoderm present along the interface of dorsal and ventral surfaces of the limb bud. Corresponding sections of *p63*^{−/−} specimens provided no evidence of a discernible AER structure. Instead, mutant limb buds had a single-layered epithelium at the distal tip, indistinguishable from the surrounding ectoderm (Fig. 2c).

Fibroblast growth factor 8 (FGF-8) can recapitulate the morphogenetic activities of the ectoderm that are necessary for the initiation and maintenance of limb outgrowth, and is a key indicator of the AER's functional integrity^{12,13}. We analysed FGF-8 expression in *p63*-mutant embryos by whole mount *in situ* hybridization. Significantly, a weak FGF-8 signal was detected on the ventral surface of *p63*^{−/−} forelimb buds at E9.5, before the formation of an AER. From E9.5 onwards, however, when the distal ventral ectoderm thickens and begins the stratification process that ultimately gives rise to an AER¹⁴, FGF-8 expression was progressively reduced in the *p63*-mutant limb bud compared with wild-type littermates (Fig. 2d). FGFs secreted by AER cells induce limb outgrowth by maintaining the underlying mesenchyme, known as the progress zone, in an undifferentiated, highly proliferative state⁹. Expression of progress-zone markers, such as *Msx1*, thus serves as an indicator for AER activity^{15,16}. AER activity was indeed reduced in the *p63*-mutant embryo, as indicated by the reduced *Msx1* expression in the progress zone (Fig. 2e).

There was a remnant of FGF-8 expression in mutant limb buds that was confined to the ventral surface, even at later stages when differential growth of the dorsal and ventral ectoderm normally results in a shift of FGF-8 to the distal tip¹⁴. One possibility, then, for the failure of AER formation in *p63*^{−/−} mice was a disruption of dorsal/ventral (D–V) patterning in the limb bud. We therefore examined the expression of *Lmx1b* (ref. 17), a dorsal mesenchyme marker, and of *Wnt7a* (refs 18, 19), a dorsal ectoderm marker, in mutant limb buds. Both *Lmx1b* and *Wnt7a* showed a more ventral spatial distribution than is seen in the wild type (Fig. 2f, and data not shown). A defined, albeit ventrally displaced D–V border was present in mutant limb buds, as judged by the absence of *Lmx1b* and *Wnt7a* in the ventral domain. Together, these results indicate that D–V patterning itself was not perturbed in the mutant embryos. Instead, they suggest that the lack of a proper AER in *p63*^{−/−} limb buds results from a failure of the ectoderm to undergo the growth and differentiation that give rise to this stratified epithelium. This inability to maintain the structural integrity of the AER, in turn, probably underlies the defects in limb morphogenesis observed in *p63*^{−/−} mice.

p63 is strongly expressed in the basal, or progenitor, cells of several epithelial structures present in the epidermis, cervix, urogenital tract, prostate and breast¹. The importance of *p63* function in epithelial differentiation was revealed by the remarkable absence of skin and related appendages in late embryonic and newborn *p63*-deficient mice (Fig. 1). Histological analysis confirmed the absence of a stratified epidermis in *p63*^{−/−} mice, as they lack the characteristic structure of basal, suprabasal and cornified layers, and associated hair follicles (Fig. 3a). In addition, tissues derived from the same stem cells that give rise to the epidermis, including mammary, sebaceous, lacrimal and salivary glands, are also absent in the *p63*^{−/−}

mouse (results not shown). Instead of an epidermis, $p63^{-/-}$ late-stage embryos and newborns retain isolated patches of disorganized epithelial cells along the exposed dermis (Fig. 3a, arrow). Further histological analysis of tissues from late-stage embryo or newborn $p63$ mutants demonstrated an absence of the normally stratified squamous epithelium on the tongue, and its replacement by one or two layers of cuboidal epithelium (Fig. 3a). Abnormalities were also observed at other sites of squamous epithelia in the mouse, including the oesophagus and proximal portion of the stomach. The oesophageal lining in $p63^{-/-}$ mice showed a pseudostratified columnar appearance, similar to that found in respiratory, but not normal oesophageal, epithelium. Periodic acid–Schiff base (PAS) staining for carbohydrates further revealed an abundance of ciliated and

goblet cells in the oesophageal epithelium of $p63^{-/-}$ mice, not found in wild-type sections (Fig. 3b). In the wild-type mouse, a stratified squamous epithelium covers the anterior portion of the stomach, whereas the distal regions are lined with a glandular epithelium. This glandular epithelium appears normal in the $p63$ mutant, but the squamous epithelium is replaced by an unusual array of cuboidal cells (Fig. 3c). A similar transfiguration was seen in the vaginal and cervical epithelium of homozygous mutants (results not shown). Finally, the transitional epithelia of the urinary bladder was thinner than normal and in some regions was missing altogether (not shown). The origin of the ectopic goblet and ciliated cells in the oesophagus, as well as that of cells replacing squamous epithelia at the cervix, anterior stomach and tongue, is unclear, but may

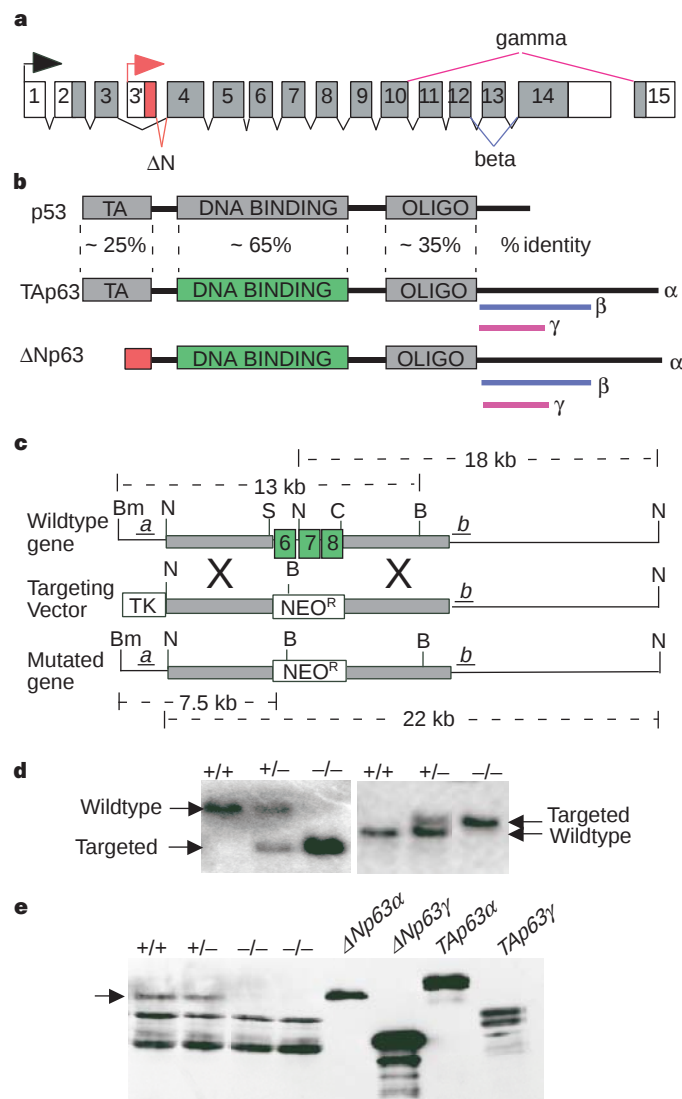
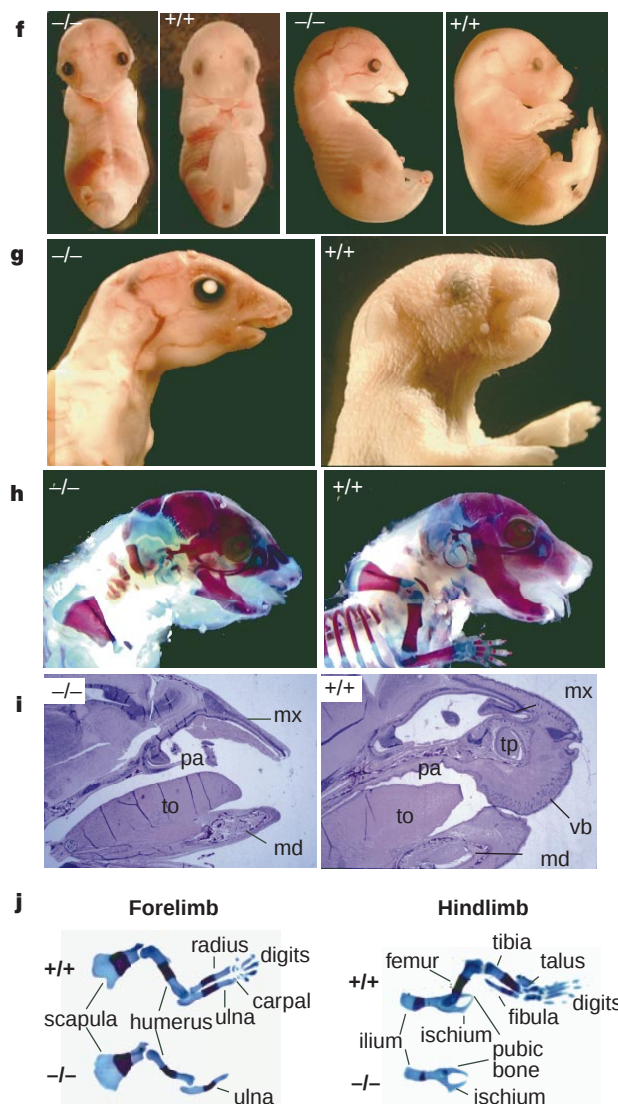


Figure 1 Disruption of the $p63$ gene in mice. **a**, Structure of murine $p63$ gene and alternative splicing events giving rise to the major $p63$ isotypes. **b**, TA- and Δ N- $p63$ isotypes derived from alternative promoters. **c**, Strategy for replacement of exons 6, 7 and 8 and flanking introns with a neomycin-resistance cassette (neo^r). Restriction sites, positions of genomic probes a and b , and the approximate sizes of hybridizing fragments of wild-type and mutant genes, are indicated (see Methods). B, *Bgl*I; Bm, *Bam*HI; C, *Cl*AI; N, *Nhe*I; S, *Spe*I. **d**, Southern-blot analysis of representative genotypes from $p63$ heterozygote crosses. **e**, Immunoblotting of whole-embryo extracts (lanes 1–4) with the 4A4 anti- $p63$ mouse monoclonal antibody. Marker lanes (5–8) are lysates of BHK cells transfected with various epitope-tagged $p63$ cDNAs, as indicated. The predominant $p63$ isotype in the embryonic extracts corresponds to Δ N- $p63\alpha$ (arrow). **f**, Front and sagittal views of



E17 control (+/+) and $p63^{-/-}$ embryos. **g**, $p63^{-/-}$ mice on postnatal day 1 (P1) have hypoplastic upper and lower jaws, and have no eyelids, whisker pads, skin and related appendages, including vibrissae, pelage follicles and hair shafts which are present on the wild-type control. **h**, Bone (Alizarin red S) and cartilage (Alcian blue) stains of P1 $p63^{-/-}$ and control mice. **i**, Haematoxylin-and-eosin (H + E) stained sagittal sections reveal a truncated maxilla (mx), mandible (md), and secondary palate (pa) and an absence of vibrissae (vb), tooth primordia (tp) and soft-tissue structures in P1 $p63^{-/-}$ mice. **j**, Bone and cartilage of fore- and hindlimbs of E15 $p63^{-/-}$ and control littermates. Homozygous mutants lack distal components of the forelimb, including the radius, carpals and digits, as well as all components of the hindlimb.

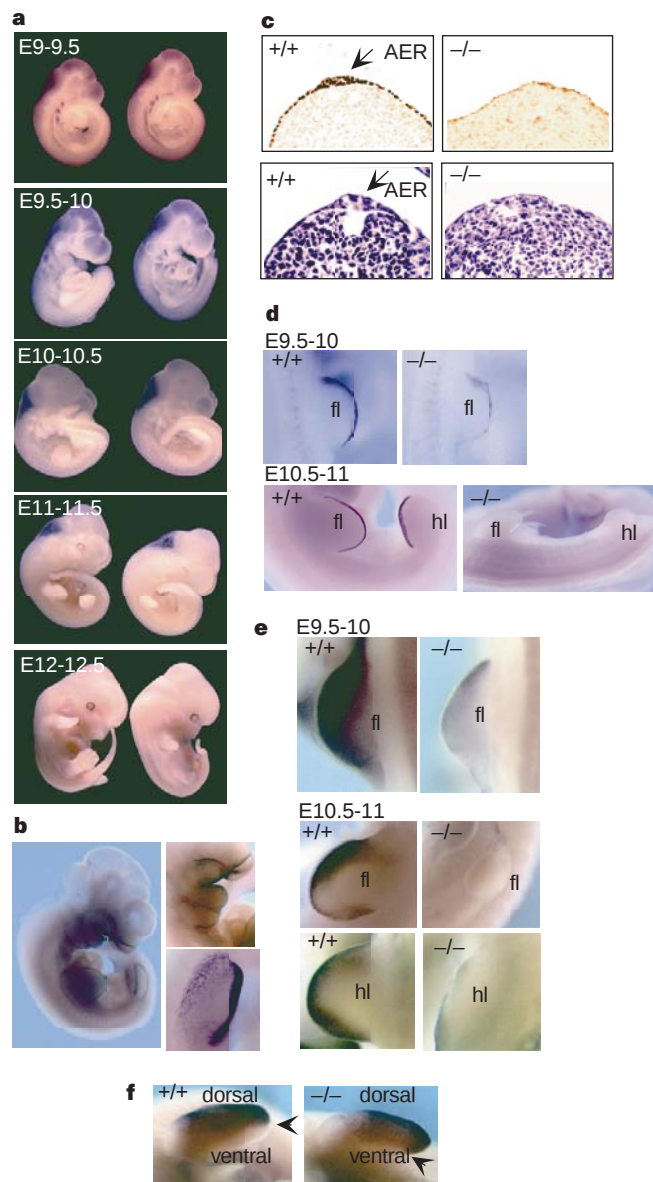


Figure 2 Defective limb development in $p63^{-/-}$ mice. **a**, Comparison of limb development in $p63^{-/-}$ (right) and control littermates (left) at embryonic days 9–12.5, showing dysmorphic and reduced limb buds in $p63^{-/-}$ embryos. **b**, Whole-mount *in situ* hybridization of a wild-type, E10.5 embryo with a $p63$ RNA probe, showing strong transcript expression in the branchial arches (upper right) and developing limb bud (lower right). **c**, Top, immunohistochemical staining with the 4A4 anti- $p63$ antibody, showing $p63$ protein expression in the apical ectodermal ridge (AER, arrow) and surrounding ectoderm of a wild-type, E11 limb bud (left) and the much weaker labelling in the corresponding $p63^{-/-}$ sections. Bottom, H + E staining of transverse sections of the limb bud at E11. Wild-type limb bud shows a multilayered ectoderm at the dorsal/ventral junction, corresponding to the AER (arrow). The AER is absent in the $p63^{-/-}$ specimen, which only shows a simple, single-layered epithelium across the limb bud. **d**, Whole-embryo *in situ* hybridization with an FGF-8 probe. At E9.5–10, compared with wild type, the FGF-8 signal in the $p63^{-/-}$ limb bud is weaker and restricted to the ventral ectoderm at both E9.5–10 and E10.5–11. fl, Forelimb bud; hl, hindlimb bud. **e**, Whole embryo *in situ* hybridizations with an $Msx-1$ probe. $Msx-1$ expression is reduced in the $p63^{-/-}$ limb bud at E9.5–10 and is almost absent at E10.5–11. **f**, Whole-embryo *in situ* hybridization with an $Lmx-1b$ probe at E10.5. $Lmx-1b$ staining is seen in the dorsal mesenchyme of both the $p63^{-/-}$ and wild-type forelimb bud. In the $p63$ mutant, however, $Lmx-1b$ shows a more ventral distribution, relative to the wild-type limb, indicating a ventral shift in the dorsoventral boundary (arrows).

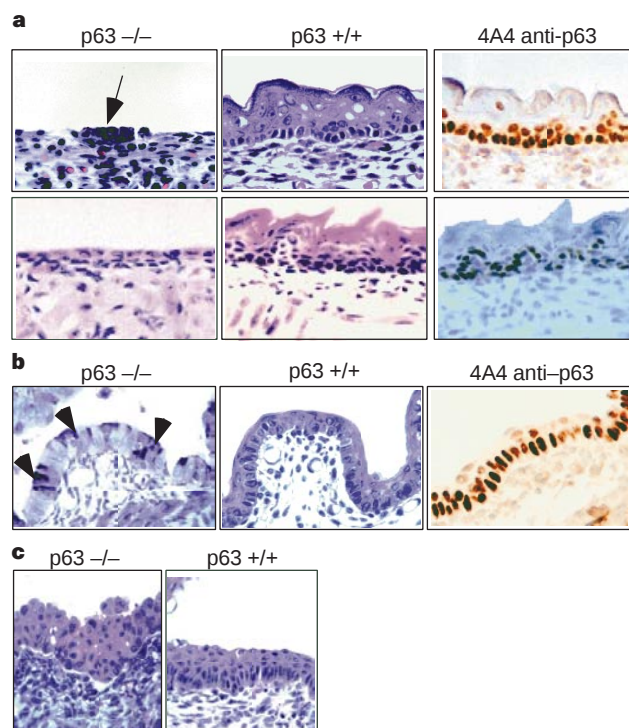


Figure 3 Defects in stratified epithelial differentiation in $p63$ -deficient mice. **a**, Left, H + E-stained sections of perinatal (E17-P1) $p63^{-/-}$ mice lacking squamous stratification in the epidermis (top) and tongue epithelium (bottom); arrow shows cellular aggregates seen along the exposed dermis. Middle, wild-type control H + E sections showing extensive stratification, with $p63$ protein expression (right) in the basal cells of these epithelia. **b**, Periodic acid-Schiff (PAS)-haematoxylin stained sections of oesophageal epithelium in $p63^{-/-}$ and control perinatal mice. Homozygous mutants lack the stratified, squamous epithelium seen in controls, and instead show an abundance of ciliated and PAS-positive goblet cells (arrows). **c**, H + E staining of the proximal stomach epithelium showing highly irregular, disorganized layers of cuboidal cells in $p63^{-/-}$ perinatal mice. Control sections show an orderly arrangement of basal cells and squamous stratification not seen in mutants.

represent a compensatory migration of surrounding epithelia or a metaplastic differentiation²⁰.

These findings highlight a strong correlation between $p63$ expression in epithelial progenitor cells and its absolute requirement in the normal development of stratified squamous epithelia. The intestinal epithelia, whose progenitor crypt cells do not stain with anti- $p63$ antibody, appear unperturbed in $p63^{-/-}$ mice (results not shown). To investigate the epithelial defects in $p63^{-/-}$ mice, we analysed the expression of squamous markers throughout embryonic skin development. Mouse embryonic skin develops from an initial, simple ectoderm that undergoes stratification beginning at E13.5 to E14. Before stratification, however, this simple epithelium initiates the expression of gene products characteristic of the basal, or progenitor, cells of stratified squamous epithelium, including the AP2 transcription factor and keratins 5 and 14 (K5, K14)²¹. Using a polyclonal antibody, we first detected K5 expression in embryonic skin of wild-type controls at E12. $p63$ -mutant embryos at this stage, however, showed no K5 staining, despite the presence of an intact, simple epithelium (Fig. 4a). At E13.5 and E17, when control embryos show increasingly strong staining for K5 and obvious stratification, $p63^{-/-}$ mutants still gave no detectable signal, concomitant with an increasingly exfoliated appearance of the epithelium (Fig. 4a, and data not shown).

The failure of the $p63^{-/-}$ primitive ectoderm to express the K5 basal marker indicated a fundamental defect in either the commitment to squamous epithelial lineages, or in the maintenance of a

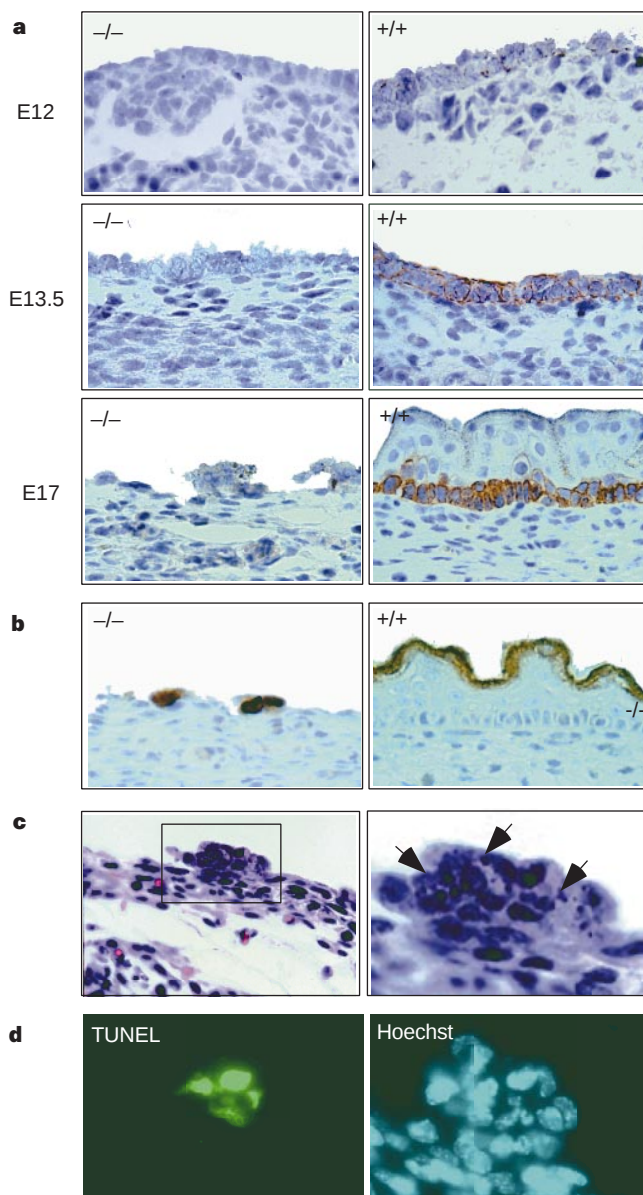


Figure 4 Expression of differentiation markers in $p63^{-/-}$ epidermis. **a**, Anti-keratin-5 (K5) antibody staining in epidermis of $p63^{-/-}$ and control littermates at various stages of embryonic development, as indicated. K5 is detected in control epidermis at E12, with increasing expression seen upon the onset of squamous stratification at $\sim$ E13.5 through late-stage embryogenesis (E17). $p63^{-/-}$ mice show no detectable K5 staining at any of these stages. **b**, Anti-loricrin antibody shows localization in the upper granular layer of wild-type E17 epidermis. Aggregates of cells along the exposed dermis of $p63^{-/-}$ littermates also stain strongly for loricrin. **c**, Left, H + E-stained section of epidermis of E17 $p63^{-/-}$ mouse showing cell aggregates; right, higher magnification of marked region showing fragmented nuclei (arrows). **d**, TUNEL assay on E17 $p63^{-/-}$ sections showing positive labelling (left) of cell aggregates along the exposed dermis; right, Hoechst DNA stain of the same section, showing hypercondensed nuclei in TUNEL-positive cells.

progenitor stem-cell population necessary for regenerative epithelia^{22–25}. To distinguish between these possibilities, we analysed the expression of loricrin, a marker for mature, terminally differentiated squamous cell populations^{26,27}. Anti-loricrin antibodies stained the upper granular layer of wild-type E17 epidermis. Cells in the abnormal clusters of epithelial cells remaining on the exposed dermis of E17 $p63^{-/-}$ embryos also showed intense loricrin expression (Fig. 4b). These epithelial cell clusters also stained positive with antibodies against two additional terminal differentiation markers,

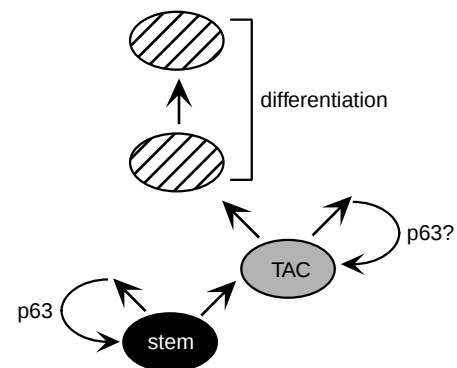


Figure 5 Model for p63 in maintaining the proliferative capacity of epithelial progenitor cells. Stem cells in the basal layer of stratified squamous epithelia express high levels of p63 and undergo asymmetric division to enable both self-renewal and progression to transient amplifying cells (TACs). TACs, which may express less p63, are also capable of limited proliferation and self-renewal, but are ultimately destined for terminal differentiation. The absence of p63 results in the failure to maintain a basal cell population, suggesting a requirement for p63 in the regenerative aspect of stem-cell division.

filaggrin and involucrin²⁷ (data not shown). Moreover, these cells contained highly pyknotic nuclei which were positive for TUNEL labelling and thus indicative of apoptosis (Fig. 4c, d). Taken together, our results indicate that the embryonic epidermis of $p63$ mutants commits to squamous lineage and undergoes a process of differentiation and cell death normally associated with epidermal tissue. Therefore, the epithelial defects in these mice are likely to involve the loss of the regenerative population of cells, and implicate p63 in functions that are related to the identity or renewal capacity of epithelial stem cells.

A common feature of the pleiotropic defects in $p63^{-/-}$ mice is the apparent failure to develop or maintain various specialized ectoderm and mesenchymal structures dependent on ectodermal signalling. The limb-bud AER, epidermis and other squamous epithelia all represent ectoderm that stratifies to form an epithelium with multiple cellular layers. Significantly, a population of progenitor stem cells along the basal layer is essential for the continued, regenerative proliferation and concomitant differentiation of these stratified epithelia^{22–25}. In culture, basal keratinocytes display differential capacities for proliferation, representative of highly regenerative stem cells, as well as 'transient amplifying' cells with more limited proliferative potential. *In vivo*, the self-renewing stem cells may act as a reserve pool for these transient amplifying cells, which then commit to an upward differentiation characteristic of stratified squamous epithelia (Fig. 5). This maturation occurs in $p63^{-/-}$ epidermis, as judged by the presence of terminal differentiation markers. Considered together with the observation that p63 is highly expressed in basal cells but is rapidly degraded upon their differentiation (ref. 1, and our unpublished observations), it is likely that p63 preserves the self-renewal capacity of progenitor cells that undergo asymmetric division to yield simultaneous regenerative and differentiated cell fates²⁸ (Fig. 5). Identifying the targets of p63 activity, as well as the mechanisms that modulate its expression during epithelial differentiation, represent important challenges. Finally, although p53-related genes have a potential for functional overlap^{1,6}, the strikingly different phenotypes of $p63^{-/-}$ and $p53^{-/-}$ mice^{29,30} highlight unique roles for p63 in development and cellular regulation. □

Methods

Targeted disruption of murine p63. A 16.5-kb clone containing the 3' end of intron 4 through to exon 10 was isolated from a 129/SvJae mouse placental genomic DNA library. To generate the targeting vector depicted in Fig. 1b, a 2.3-kb *SpeI*–*Clal* fragment, corresponding to exons 6, 7 and 8 and flanking

introns, was deleted and replaced with the neomycin-resistance (*neo^r*) gene driven by the mouse phosphoglycerol kinase (PGK1) promoter, in reverse orientation to the *p63* gene. To select against random insertional events, a thymidine kinase (TK) gene under the regulation of MC1 promoter⁷ was incorporated at the 5' end of the targeting vector. The targeting vector contained ~5 kb of homology on each side of the PGK-*neo^r* gene. The vector was linearized using a unique *NofI* site and electroporated into the J1 ES cell line. G418 and FIAU-resistant ES cell colonies were picked, expanded and screened using probe *a* to hybridize genomic DNA digested with *Bgl*I and *Bam*HI. Probe *a* hybridizes to a DNA fragment of ~13.5 kb corresponding to the wild-type allele, and to a 7.5-kb fragment resulting from the targeted allele. Positive clones were screened with probe *b*, which hybridizes to ~18-kb wild-type fragment and ~22-kb targeted allele upon *Nhe*I digestion. Two distinct ES cell lines heterozygous for the disrupted allele were microinjected into blastocysts from B57BL/6 and BALB/c mice, causing germline transmission of the *p63* mutation. Mice heterozygous for the mutant *p63* allele were interbred, and the genotype of progeny determined by Southern blotting as described. Harvard Medical School is an AAALAC accredited institution and the mice were cared for in accordance with institutional guidelines.

Histology and immunohistochemistry. Specimens were fixed in Bouin's fluid, 10% buffered formalin, or 4% paraformaldehyde in PBS, embedded in paraffin, and sectioned at 6 mm. Histological analysis was done on sections stained with haematoxylin and eosin. *p63* staining was done with the 4A4 anti-*p63* monoclonal antibody¹; K-5, loricrin, involucrin, and flaggrin staining was with rabbit polyclonal antibodies^{24,27} (provided by P. Dotto and H. Green).

In situ hybridization. Non-radioactive whole-mount *in situ* hybridization of mouse embryos was done essentially as described¹⁸. Probes used were: Δ Np63 γ (ref. 1), Wnt7a (ref. 19), FGF-8 (ref. 12), and Lmx1b (ref. 17).

Received 21 January; accepted 17 March 1999.

- Yang, A. *et al.* *p63*, a *p53* homolog at 3q27-29, encodes multiple products with transactivating, death-inducing, and dominant-negative activities. *Mol. Cell* **2**, 305–316 (1998).
- Augustin, M., Bamberger, C., Paul, D. & Schmale, H. Cloning and chromosomal mapping of the human *p53*-related KET gene to chromosome 3q27 and its murine homolog Ket to mouse chromosome 16. *Mamm. Genome* **11**, 899–902 (1998).
- Osada, M. *et al.* Cloning and functional analysis of human *p51*, which structurally and functionally resembles *p53*. *Nature Med.* **4**, 839–843 (1998).
- Trink, B. *et al.* A new human *p53* homologue. *Nature Med.* **4**, 747–748 (1998).
- Ko, L. J. & Prives, C. *p53*: puzzle and paradigm. *Genes Dev.* **10**, 1054–1072 (1996).
- Kaghad, M. *et al.* Monoallelically expressed gene related to *p53* at 1p36, a region frequently deleted in neuroblastoma and other human cancers. *Cell* **90**, 809–819 (1997).
- Mansour, S. L., Thomas, K. R. & Capocchi, M. R. Disruption of the proto-oncogene *int-2* in mouse embryo-derived stem cells: a general strategy for targeting mutations to non-selectable genes. *Nature* **336**, 348–352 (1988).
- Tickle, C. Vertebrate limb development. *Curr. Opin. Genet. Dev.* **5**, 478–484 (1995).
- Johnson, R. L. & Tabin, C. J. Molecular models for vertebrate limb development. *Cell* **90**, 979–990 (1997).
- Kelley, R. O. & Fallon, J. F. Ultrastructural analysis of the apical ectodermal ridge during vertebrate limb morphogenesis. I. The human forelimb with special reference to gap junctions. *Dev. Biol.* **51**, 241–256 (1976).
- Francis-West, P., Ladherr, R., Barlow, A. & Graveson, A. Signalling interactions during facial development. *Mech. Dev.* **75**, 3–28 (1998).
- Crossley, P. H., Minowada, G., MacArthur, C. A. & Martin, G. R. Roles for FGF-8 in the induction, initiation, and maintenance of chick limb development. *Cell* **84**, 127–136 (1996).
- Vogel, A., Rodriguez, C. & Izpisua-Belamonte, J. C. Involvement of FGF-8 in initiation, outgrowth, and patterning of the vertebrate limb. *Development* **122**, 1737–1750 (1996).
- Loomis, C. A., Kimmel, R. A., Tong, C. X., Michaud, J. & Joyner, A. L. Analysis of the genetic pathway leading to formation of ectopic apical ectodermal ridges in mouse *Engrailed-1* mutant limbs. *Development* **125**, 1137–1148 (1998).
- Roberts, B., Lyons, G., Simandl, B. K., Kuroiwa, A. & Buckingham, M. The apical ectodermal ridge regulates Hox-7 and Hox-8 gene expression in developing chick limb buds. *Genes Dev.* **5**, 2362–2374 (1991).
- Wang, Y. & Sassoon, D. Ectoderm-mesenchyme and mesenchyme-mesenchyme interactions regulate *Mx-1* expression and cellular differentiation in the murine limb bud. *Dev. Biol.* **168**, 374–382 (1995).
- Chen, H. *et al.* Limb and kidney defects in Lmx1b mutant mice suggest an involvement of LMX1B in human nail patella syndrome. *Nature Genet.* **19**, 51–55 (1998).
- Riddle, R. D. *et al.* Induction of the LIM homeobox gene Lmx1 by WNT7a establishes dorsoventral pattern in the vertebrate limb. *Cell* **83**, 631–640 (1995).
- Parr, B. A. & McMahon, A. P. Dorsalizing signal Wnt-7a required for normal polarity of D–V and A–P axes of mouse limb. *Nature* **374**, 350–353 (1995).
- Pellegrini, G. *et al.* Location and clonal analysis of stem cells and their differentiated progeny in the human ocular surface. *J. Cell Biol.* (in the press).
- Byrne, C., Tainsky, M. & Fuchs, E. Programming gene expression in developing epidermis. *Development* **120**, 2369–2383 (1994).
- Rheinwald, J. G. & Green, H. Serial cultivation of strains of human epidermal keratinocytes: the formation of keratinizing colonies from single cells. *Cell* **6**, 331–344 (1975).
- Barrandon, Y. & Green, H. Three clonal types of keratinocytes with different capacities for multiplication. *Proc. Natl Acad. Sci. USA* **84**, 2302–2306 (1987).
- Barrandon, Y. The epidermal stem cell: an overview. *Dev. Biol.* **4**, 209–215 (1993).
- Watt, F. M. Epidermal stem cells: markers, patterning and the control of stem cell fate. *Phil. Trans. R. Soc. Lond. B* **358**, 553–560 (1991).
- Greer, J. & Roop, D. Loricrin, a major keratinocyte envelope protein, is expressed late in development. *J. Invest. Derm.* **96**, 553–560 (1991).

- Calautti, E., Missero, C., Stein, P. L., Ezzell, R. M. & Dotto, G. P. Fyn tyrosine kinase is involved in keratinocyte differentiation control. *Genes Dev.* **9**, 2279–2291 (1995).
- Jan, Y. N. & Jan, L. Y. Asymmetric cell division. *Nature* **392**, 775–778 (1998).
- Donehower, L. A. *et al.* Mice deficient for *p53* are developmentally normal but susceptible to spontaneous tumours. *Nature* **356**, 215–221 (1992).
- Jacks, T. *et al.* Tumor spectrum analysis in *p53* mutant mice. *Curr. Biol.* **4**, 1–7 (1994).

Acknowledgements. We thank H. Green, M. Kirschner, P. Dotto, B. Olsen, N. Fukai, T. Rapoport, B. Quade, J. Glickman, P. Ferrara and T. Bestor for their support and for discussion; H. Green, G. Martin, A. McMahon, R. Johnson, A. Joyner and P. Dotto for reagents; L. Du for blastocyst injections; H. Liu for the 129 mouse genomic library; and J. Williams for histology preparations. This work was supported by grants from the NIH to R.B., C.T., A.S. and F.M.

Correspondence and requests for materials should be addressed to F.McK. (e-mail: frank_mckeeon@hms.harvard.edu).

MCP-1 and IL-8 trigger firm adhesion of monocytes to vascular endothelium under flow conditions

Robert E. Gerszten^{*†}, Eduardo A. Garcia-Zepeda^{†‡}, Yaw-Chyn Lim^{†§}, Masayuki Yoshida^{||}, Han A. Ding[§], Michael A. Gimbrone Jr^{†§}, Andrew D. Luster^{†‡}, Francis W. Luscinskas^{†§} & Anthony Rosenzweig^{*†}

^{*} The Cardiovascular Research Center and Cardiology Division, Massachusetts General Hospital, Boston, Massachusetts 02114, USA

[†] Harvard Medical School, Boston, Massachusetts 02114, USA

[‡] Infectious Disease Unit, Massachusetts General Hospital, Boston, Massachusetts 02114, USA

[§] Vascular Research Division, Department of Pathology, Brigham and Women's Hospital, Boston, Massachusetts 02115, USA

^{||} Molecular Genetics, Medical Research Institute, Tokyo Medical and Dental University, Tokyo, Japan

Monocytes contribute to the development of atherosclerotic lesions in mouse models^{1–3}. The chemoattractant proteins (chemokines), monocyte chemoattractant protein-1 (MCP-1) and interleukin-8 (IL-8), are found in human atheroma^{4,5}, and mice lacking receptors for these chemokines are less susceptible to atherosclerosis and have fewer monocytes in vascular lesions^{6,7}. Although MCP-1 has a powerful effect on monocytes, IL-8 is thought to act predominantly on neutrophils and it is unclear how it could recruit monocytes^{6,8}. Here we investigate the ability of chemokines to control the interaction of monocytes under flow conditions with vascular endothelium that has been transduced to express specific leukocyte-adherence receptors. We find that MCP-1 and IL-8 can each rapidly cause rolling monocytes to adhere firmly onto monolayers expressing E-selectin, whereas related chemokines do not. These effects do not correlate with either the induction of a calcium transient or chemotaxis. We conclude that chemokines are important modulators of monocyte–endothelial interactions under flow conditions. Moreover, our finding that IL-8 is a powerful trigger for firm adhesion of monocytes to vascular endothelium reveals an unexpected role for this chemokine in monocyte recruitment.

To investigate the mechanisms of monocyte adhesion, we used recombinant adenoviruses containing specific endothelial adhesion molecules to transduce human umbilical vein endothelial cells (HUVEC), a well characterized model of vascular endothelium. The initial tethering of monocytes to vascular endothelium can be mediated by vascular-cell adhesion molecule-1 (VCAM-1), but we found that VCAM-1 alone was unable to support firm monocyte adhesion under flow as well as cytokine-activated endothelium could⁹, indicating that other signals expressed by activated endothelial cells were enhancing monocyte recruitment⁹. At relatively higher shear (≥ 1.5 dynes cm^{–2}), the initial rate of monocyte attachment to

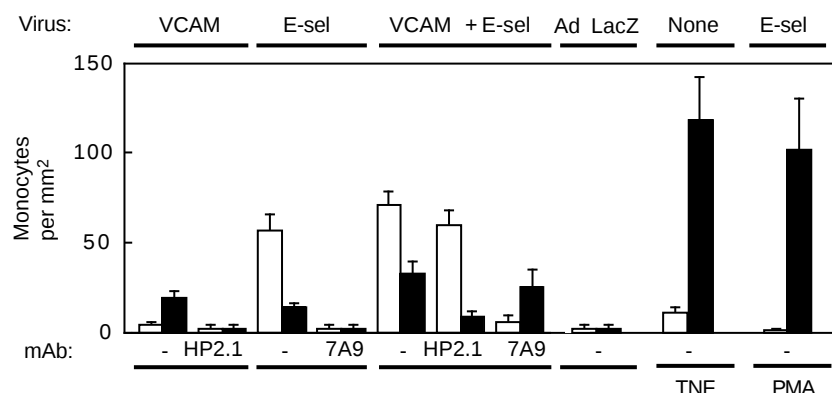


Figure 1 Monocyte attachment to adenovirally transduced or cytokine-activated HUVEC. Endothelial monolayers were transduced with AdVCAM-1, AdE-sel, or both, or with AdRSVβgal(LacZ), or were stimulated with TNF for 4 h immediately before the experiment. 48 h after infection, the interaction of monocytes with these monolayers was studied at 2.0 dynes cm⁻². The phenotype of sustained monocyte–endothelial interactions was characterized as either ‘rolling’ (white bars) or ‘firm adhesion’ (black bars)⁹. AdVCAM-1 supported predominantly monocyte firm adhesion, whereas AdE-sel supported predominantly monocyte rolling. Compared to VCAM-1 alone, co-expression of E-selectin with VCAM-1 greatly increased monocyte rolling ($n = 7$, $P < 0.001$) with only a modest increase in firm adhesion ($n = 7$, $P < 0.02$). In contrast, TNF-activated HUVEC supported much more monocyte firm adhesion ($n = 6$, $P < 0.001$) and less rolling ($n = 6$,

$P < 0.001$) than did doubly transduced HUVEC. Monoclonal antibodies (HP2.1) against the VCAM-1 counterligand ($\alpha 4$ -integrin) effectively blocked monocyte adhesion to AdVCAM-transduced HUVEC ($n = 3$, $P < 0.001$). HP2.1 also blocked monocyte firm adhesion to doubly transduced HUVEC ($n = 3$, $P < 0.01$), resulting in predominantly monocyte rolling. A monoclonal antibody (mAb) (7A9) directed at E-selectin blocked all monocyte interactions with AdE-sel-transduced HUVEC ($n = 3$, $P < 0.001$) and restored a pattern similar to that seen with the expression of VCAM-1 alone. Activation of monocytes with PMA (100 ng ml⁻¹) increased monocyte firm adhesion to E-selectin-transduced monolayers to levels comparable to those seen in TNF-activated HUVEC ($n = 3$, $P = \text{NS}$). Data shown ($\pm$ s.e.m.) are pooled from nine experiments, with each condition tested the indicated number of times.

HUVEC activated by tumour-necrosis factor (TNF) significantly exceeded the attachment to VCAM-1-transduced HUVEC. Co-expression of E-selectin with VCAM-1 significantly increased primary monocyte–endothelial interactions (data not shown), but these interactions manifested mainly as monocyte rolling, rather than as firm adhesion (Fig. 1). Thus, vascular endothelium

expressing E-selectin and VCAM-1 at levels comparable to those in cytokine-activated HUVEC supports more monocyte rolling and less adhesion than do TNF-activated HUVEC. In contrast, activation of monocytes with phorbol myristate acetate (PMA; 100 ng ml⁻² for 5 min at 37 °C) enhances firm adhesion even to monolayers transduced with E-selectin alone, to a similar extent to cytokine-activated endothelium (Fig. 1). These results suggest that leukocyte activation is important for monocyte recruitment, so we have investigated the ability of chemokines, as pathophysiologically relevant leukocyte activators expressed by cytokine-treated HUVEC, to mediate the conversion of monocyte rolling to firm adhesion.

Chemokines were added to monocytes being perfused over E-selectin-transduced monolayers, as a model of selectin-mediated monocyte rolling. Monocyte rolling and firm adhesion were quantified for each HUVEC monolayer one minute before and one minute after addition of the chemokine. Recombinant MCP-1 (a CC chemokine) at concentrations ≥ 250 pM effectively transformed virtually all monocyte rolling to firm adhesion (Fig. 2), whereas MCP-4, a CC chemokine that induces monocyte chemotaxis and Ca²⁺ transients at doses as low as 2.5 nM (ref. 10), only moderately increased monocyte firm adhesion, even at 100 nM. Surprisingly, IL-8 (a CXC chemokine) at concentrations ≥ 2 nM induced monocyte firm adhesion comparable to that seen after MCP-1 treatment (Fig. 2), whereas epithelial-derived neutrophil-activating protein (ENA)-78 (Fig. 2) and other CXC chemokines (IP-10 and PF-4; data not shown) had little or no effect on E-selectin-dependent monocyte rolling. IL-8-induced monocyte arrest occurred with recombinant chemokine from two different commercial sources and was prevented by monoclonal antibody against IL-8 (data not shown). Several factors indicate that the effect on adhesion was not due to the *de novo* recruitment of other leukocytes such as neutrophils, which might contaminate the monocyte preparations. First, the purity of the starting population was confirmed by both flow cytometry and histochemical staining. Second, histochemically stained coverslips after IL-8 or MCP-1 treatment were reviewed by clinical haematologists blinded to the treatment protocol: in each case, adherent leukocytes were a homogeneous population with the

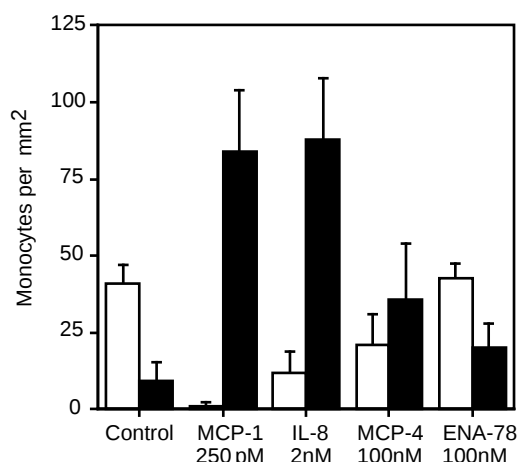


Figure 2 MCP-1 and IL-8 induce arrest of monocytes. Endothelial monolayers were transduced with AdE-sel and cultured for 48 h. Monocytes were perfused at a shear stress of 2.0 dynes cm⁻². The indicated chemokine was added to the monocyte reservoir (room temperature) and monocyte rolling (white bars) and firm adhesion (black bars) were quantified for each coverslip 1 min before and 1 min after addition of the chemokine. MCP-1 (250 pM; $n = 3$) and IL-8 (2 nM; $n = 7$) converted virtually all monocyte rolling to firm adhesion within 1 min ($P < 0.001$ for firm adhesion compared with before chemokine addition). In contrast, MCP-4 induced at most a modest increase in monocyte firm adhesion at doses up to 100 nM ($n = 5$, $P < 0.01$ compared with IL-8 or MCP-1). ENA-78 did not significantly affect monocyte rolling ($n = 5$, $P = \text{NS}$). Data ($\pm$ s.e.m.) are pooled from seven experiments, with each condition tested the indicated number of times.

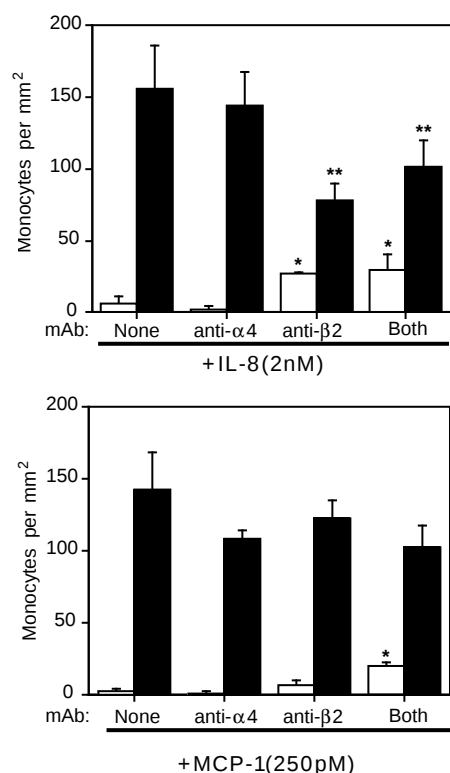


Figure 3 MCP-1 and IL-8 augment arrest by using different integrins. HUVEC monolayers were infected with AdE-sel and cultured for 48 h. Monocytes were incubated with the indicated mAb for 10 min at 4°C immediately before the experiment. IL-8-induced monocyte arrest (black bars) was significantly blocked by an mAb (TS1/18) to $\beta 2$ integrin ($P < 0.01$ compared with no mAb), with restoration of the rolling phenotype (white bars, $P < 0.001$), whereas an mAb to $\alpha 4$ integrin (HP2/1) had no effect on IL-8-induced firm adhesion (top). In contrast, mAbs to $\alpha 4$ or $\beta 2$ integrin alone did not affect MCP-1-induced monocyte adhesion. In combination, these mAbs could restore some monocyte rolling after MCP-1 treatment ($P < 0.05$), although no significant decrease in overall adhesion was noted (bottom). Binding antibody to class I MHC (W6/32) did not affect adhesion either (data not shown). Representative data are one of four experiments. Single asterisk indicates a significant increase in monocyte rolling. Double asterisks indicate a significant decrease in monocyte firm adhesion.

morphology of monocytes (data not shown). Finally, we monitored the effect of the chemokine on rolling leukocytes by videomicroscopy. The firm leukocyte adhesion seen after IL-8 or MCP-1 treatment reflected the arrest of virtually all rolling leukocytes, rather than recruitment of a different population from the flow stream. Thus, both IL-8 and MCP-1 arrested monocytes interacting with E-selectin-expressing vascular endothelium.

To investigate the adhesive mechanisms contributing to monocyte arrest in this system, we used function-blocking monoclonal antibodies to the $\alpha 4$ and $\beta 2$ leukocyte integrins. A number of counterligands for these integrins are constitutively expressed by vascular endothelium, including intercellular adhesion molecule-1 and fibronectin¹¹. IL-8-induced monocyte arrest was significantly

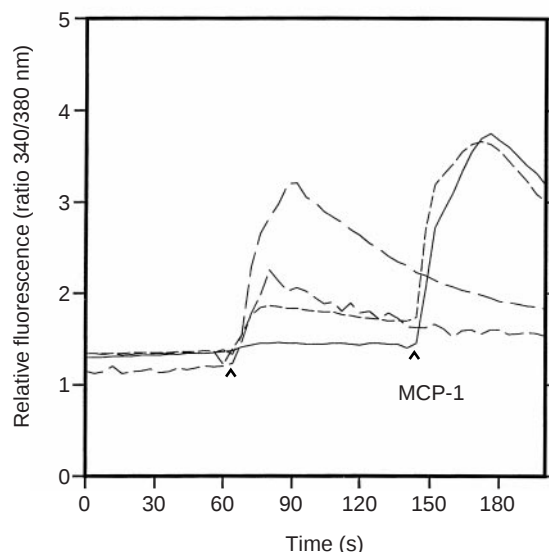


Figure 4 Chemokine-induced calcium flux in human monocytes. Fura-2-loaded monocytes were exposed to MCP-1 (2 nM; long-dash line), MCP-4 (2.5 nM; medium-dash line), IL-8 (10 nM; short-dash line), or ENA-78 (100 nM; solid line) at the time indicated (left caret). The data are presented as the relative ratio of fluorescence at 340 and 380 nm. MCP-1 induced a strong response; MCP-4 and IL-8 induced smaller calcium transients. ENA-78 did not elicit calcium flux at doses of up to 100 nM. A subsequent strong response to MCP-1 was recorded in the ENA-78 and IL-8-treated samples (right caret).

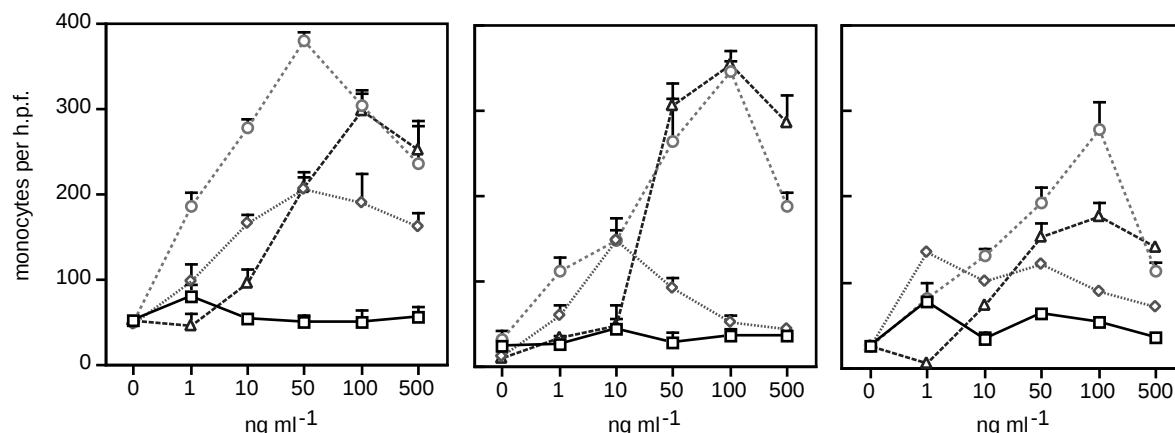


Figure 5 Chemotaxis of elutriated monocytes. Purified monocytes in a modified Boyden chamber were exposed to MCP-1 (circles), MCP-4 (triangles), IL-8 (diamonds) or ENA-78 (squares) at the indicated concentrations. After 90 min, the cells that migrated through the membrane were counted in three high-power fields (h.p.f.) from each duplicate well. Data ($\pm$ s.d.) are shown from three

independent chemotaxis experiments using monocytes elutriated from different donors, done in parallel with flow experiments. In each experiment, IL-8 induced significantly more monocyte chemotaxis than did ENA-78 ($P < 0.0001$ at 10 and 50 ng ml⁻¹).

blocked by a monoclonal antibody to $\beta 2$ integrin (TS1/18), with restoration of the rolling phenotype (Fig. 3, top). A monoclonal antibody to $\alpha 4$ integrin (HP2/1) had no effect on IL-8-induced firm adhesion. In contrast, neither of these antibodies alone affected MCP-1-induced monocyte adhesion, but in combination they restored some monocyte rolling despite MCP-1 treatment (Fig. 3, bottom). These results indicate that, although IL-8 and MCP-1 both trigger the firm adhesion of monocytes, the contribution of specific integrin adhesion receptors to the induced adhesion differs for these

two agonists, perhaps because they use different signalling mechanisms. There is early activation of $\beta 1$ integrins ($\alpha 4\beta 1$ and $\alpha 5\beta 1$) in both monocytes¹² and T lymphocytes¹³ after MCP-1 treatment, which may have important implications for transendothelial chemotaxis. In contrast, we found that function-blocking monoclonal antibodies to $\alpha 4$ had little effect under flow conditions, which probably reflects the absence of the principal endothelial $\beta 1$ counterligand, VCAM-1, as well as differences in the monocyte-endothelial interaction.

Given that the CXC chemokine IL-8, but not the CC chemokine

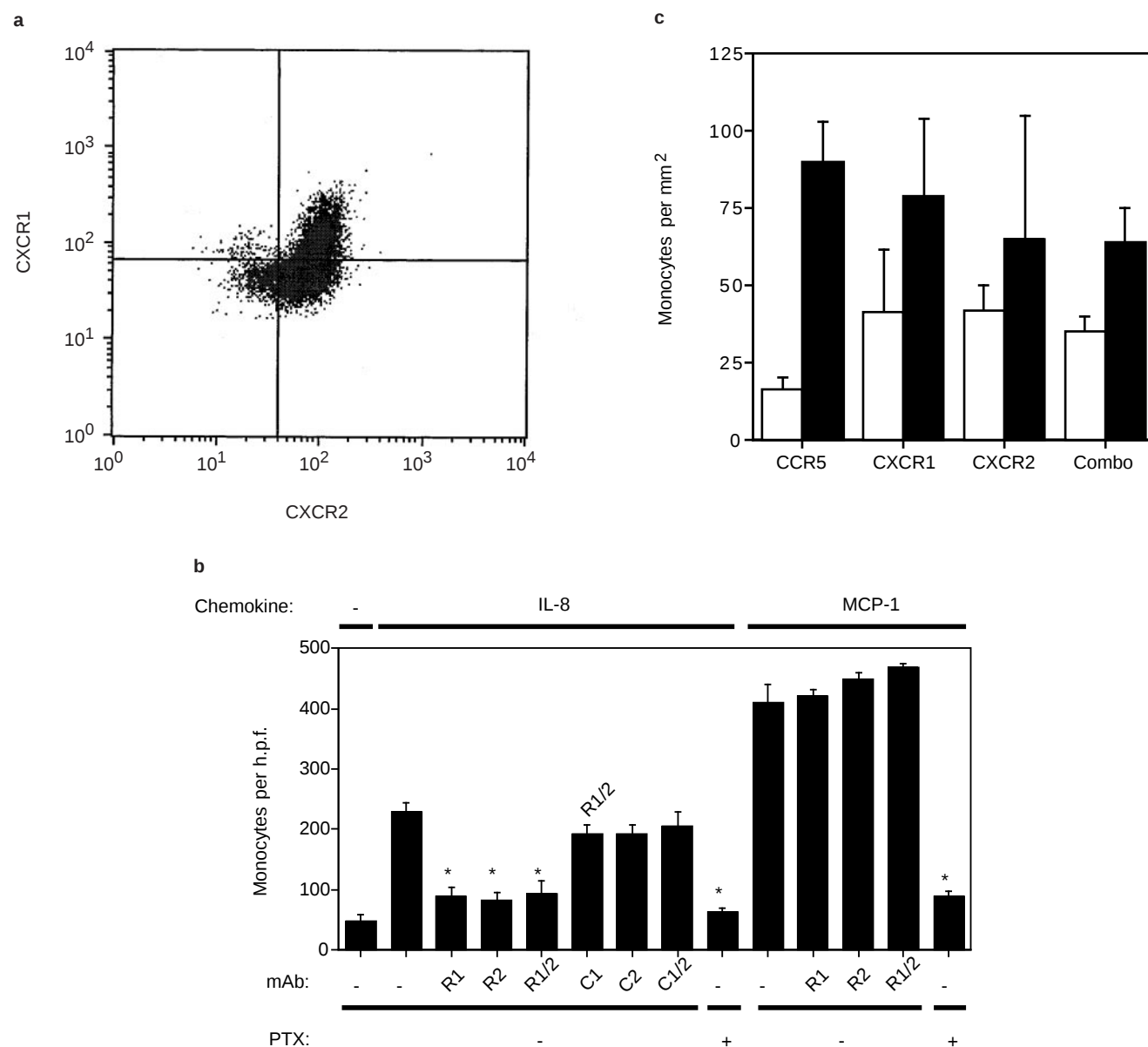


Figure 6 Monocytes express functional IL-8 receptors. **a**, Surface expression of CXCR1 and CXCR2 on elutriated monocytes. Monocytes were subjected to flow cytometry with fluorescently labelled mAb to CD14 (APC), CXCR1 (FITC) and CXCR2 (PE) (see Methods). Surface expression of CXCR1 and CXCR2 on CD14-positive cells is shown from one of four independent experiments. **b**, mAb to CXCR1 and CXCR2 inhibit IL-8 induced monocyte chemotaxis. Purified monocytes were incubated with the indicated mAb and then exposed to MCP-1 or IL-8 (both at 50 ng ml⁻¹) in a modified Boyden chamber. After 90 min, the cells that migrated through the membrane were counted in three high-power fields from each duplicate well. Representative data are from one of three independent experiments. In each experiment, treatment with mAb to CXCR1 (R1), CXCR2 (R2) or both (R1/2) significantly reduced monocyte chemotaxis to IL-8 but not to MCP-1 ($P < 0.001$ compared with IL-8 alone or treatment with isotype-matched

control mAb, C1 for IgG₁ and C2 for IgG_{2A}). Pretreatment of monocytes with pertussis toxin (PTX: 250 ng ml⁻¹ for 2 h at 37°C) blocked monocyte chemotaxis to both IL-8 and MCP-1 ($P < 0.001$ compared with chemokine alone). **c**, mAbs to CXCR1 and CXCR2 restored monocyte rolling after IL-8 treatment. HUVEC monolayers were infected with AdE-sel as above and cultured for 48 h. Monocytes were incubated with the indicated mAb for 10 min at 4°C immediately before the experiment. Monocyte rolling (white bars) and firm adhesion (black bars) were quantified 1 min after IL-8 (2 nM) addition. Representative data are from one of three experiments. Simultaneous treatment with mAbs raised against CXCR1 and CXCR2 (combo) significantly preserved the rolling phenotype as compared with control mAb ($P < 0.05$ for the experiment shown and for pooled data from all three experiments).

MCP-4, could trigger firm adhesion of monocytes at nanomolar concentrations under laminar flow, we tested whether these chemokines could induce calcium flux in elutriated monocytes. MCP-1 (2 nM) and MCP-4 (2.5 nM) both stimulate a large and rapid rise in intracellular calcium in monocytes (Fig. 4)¹⁰; IL-8 produces a small dose-dependent calcium flux in monocytes (Fig. 4), consistent with previous reports⁸. Therefore, the ability of chemokines to trigger adhesion of monocytes under flow to E-selectin-transduced endothelium does not correlate with their ability to stimulate a rise in intracellular calcium in monocytes. Moreover, the concentrations of MCP-1 and IL-8 that stimulate calcium transients in monocytes were consistently higher than those required to trigger firm adhesion of monocytes. Our assay evidently reveals an extremely sensitive biological effect.

We tested whether IL-8 could function as a monocyte agonist in other physiologically relevant models by chemotaxis assays. Although donor variability was evident, IL-8 consistently induced significant monocyte chemotaxis with a smaller maximal effect than that seen with MCP-1 or -4 (Fig. 5). In all cases, histochemical staining confirmed that the transmigrated cells were monocytes (data not shown). Moreover, checkerboard analysis performed in parallel in two experiments demonstrated that IL-8 induced monocyte chemotaxis rather than simple chemokinesis (data not shown). Although IL-8-induced monocyte chemotaxis has not been detected before^{14,15}, we used elutriated monocytes rather than cells isolated by adherence to substrate, which may alter the functional phenotype of monocytes¹⁶. As MCP-4 is more potent than IL-8 in inducing both a calcium transient and monocyte chemotaxis (Figs 4 and 5) but is less effective at converting monocyte rolling to firm adhesion (Fig. 2), these effects are presumably mediated by divergent intracellular signalling pathways.

IL-8 binds to two known receptors in humans, CXCR1 and CXCR2 (ref. 17), although only a homologue of CXCR2 has been found in mice⁶. One or both receptors were present on the surface of most of the elutriated monocytes used here (Fig. 6a). In three out of four donors, CXCR2 expression was more prevalent than CXCR1 expression, consistent with previous reports¹⁶, although the relative prevalence varied widely among monocyte donors. CXCR1 was expressed on 23–90% of the elutriated monocytes, and CXCR2 was expressed on 22–93%. Pretreatment of monocytes with monoclonal antibodies raised against either of these receptors significantly reduced monocyte chemotaxis to IL-8 but not to MCP-1 (Fig. 6b). Simultaneous treatment with both antibodies partially restored monocyte rolling after IL-8 treatment (Fig. 6c). In combination, these antibodies also inhibited IL-8-induced calcium transients (data not shown). As further confirmation of receptor-mediated signalling, pertussis toxin also blocked IL-8-induced chemotaxis (Fig. 6b); however, the same experiment could not be done for adhesion under flow because pertussis toxin activated the monocytes, as shown by loss of surface L-selectin and increased β 1 integrin activation epitopes on flow cytometry (data not shown), and it altered the baseline phenotype of monocyte–endothelial interaction. Nevertheless, the cumulative data indicate that the IL-8 receptors on monocytes¹⁶ are functional and account for the observed phenomena. ENA-78, which binds CXCR2, did not induce significant monocyte chemotaxis (Fig. 5) or firm adhesion (Fig. 2). These results are consistent with a model in which both receptors are necessary but neither alone is sufficient for the IL-8 effects, perhaps because the receptors interact cooperatively (as reported for other members of this receptor family¹⁸) or because their downstream signalling pathways may converge.

Our results indicate that specific chemokines are essential for monocyte recruitment not only as chemoattractants, but also by translating initial monocyte tethering into firm adhesion through activation of leukocyte integrin. The *in vitro* model presented here provides a sensitive system for investigating this control by chemokines and reveals an important biological effect that is not predicted

by results of simpler *in vitro* assays, such as measurement of calcium transients or chemotaxis. The unexpected finding that the CXC chemokine IL-8 can trigger monocyte firm adhesion to vascular endothelium indicates a potential role for this chemokine in monocyte recruitment and reinforces the biological complexity of the chemokine family. Identification of the vascular signals responsible for monocyte recruitment could provide insight into the cellular and molecular mechanisms of atherogenesis and suggest new targets for therapeutic intervention. □

Methods

Materials. RPMI 1640, DMEM and DPBS with or without Ca^{2+} and Mg^{2+} were purchased from BioWhittaker; HSA was from Baxter Healthcare; FBS was from Hyclone; recombinant hTNF- α was from Biogen. Recombinant chemokines were generously provided by PeproTech or purchased from R&D Systems. Pertussis toxin was purchased from Calbiochem.

Cell culture. HEK 293 cells were obtained from the American Type Culture Collection and cultured in DMEM supplemented with 10% FBS as described⁹. HUVEC were isolated and cultured as described¹⁹ in M199 with 20% FBS, endothelial-cell growth factor (25 $\mu\text{g ml}^{-1}$; Biomedical Technologies), porcine intestinal heparin (50 $\mu\text{g ml}^{-1}$; Sigma), and antibiotics; after infection with adenoviral vectors, HUVEC were cultured in 10% FBS. For stimulation of HUVEC, recombinant hTNF- α (200 U ml^{-1}) was added as indicated. For experimental use in the flow-plate apparatus, HUVEC (passage 1–2) were plated at confluence on 25-mm fibronectin-coated glass coverslips as described¹⁹; infection of HUVEC has been described⁹. On the day of the flow adhesion assay, a fluorescence immunoassay was done on HUVEC infected in parallel to determine transgene expression and to rule out nonspecific activation of the endothelial monolayer⁹.

Recombinant adenoviruses. Three recombinant type-5 adenoviruses (Ad) were used in these studies: AdRSV β -gal, AdVCAM-1, and AdE-sel. All three viruses have a similar backbone, containing E1/E3 deletions. AdRSV β -gal was a gift from D. A. Dichek (Gladstone Institute for Cardiovascular Diseases). The construction of AdVCAM-1 and AdE-sel has been described^{20,21}. Large-scale production of adenovirus and determination of viral titre was done as described⁹. Only one viral stock of each construct was used. Stock titre was 10^{10} PFU ml^{-1} for both vectors, with a particle per PFU ratio of $\sim 10^2$.

Monoclonal antibodies. The following mAb were used as purified IgG and have been described: 7A9 (ref. 22) and H4/18 (to human E-selectin²³), Hu5/3 (to human ICAM-1; ref. 24), E1/6 (which recognizes both human and rabbit VCAM-1; ref. 25), HP2.1 (to α_4 -integrin; Immunotech), 5A12 (to CXCR1; Pharmingen), 6C6 (to CXCR2; Pharmingen), 2D7 (to CCR5; Leukosite), IgG₁ and IgG_{2A} nonbinding control mAb (Pharmingen).

Flow cytometry. Monocytes were washed once with RPMI/5% FCS and incubated with the indicated fluorescently tagged primary antibodies for 30–60 min on ice, washed twice with RPMI/5% FCS, and fixed with 1% formaldehyde. An isotype-matched, fluorescently labelled nonbinding antibody was included as a control. Fluorescence was then analysed using a Becton–Dickinson FACS set to detect fluorescence, forward scatter and size.

Leukocyte isolation. Human monocytes were purified from single-donor human platelet pheresis residues by Ficoll–Hypaque density-gradient centrifugation at 15 °C (LSM; Organon Teknica), followed by counterflow centrifugation elutriation as described²⁶. Monocyte suspensions were >91% pure, with 6–8% lymphocyte, 2% granulocyte and essentially no platelet contamination, as determined by light scatter and cell-surface antigen analysis²⁶.

Adhesion assays under flow. The parallel-plate flow chamber and assay protocols have been described^{19,27}. For endothelial-blocking experiments, HUVEC monolayers were incubated immediately before the assay with culture medium containing the indicated mAb or culture medium alone for 30 min at 37 °C. For monocyte blocking, leukocytes were incubated with the indicated mAb for 10 min at 4 °C and diluted with perfusion medium to 10^6 cells per ml. Where indicated, chemokines were added to the monocyte reservoir (room temperature). Monocyte rolling and firm adhesion were quantified for each coverslip¹⁹ 1 min before and 1 min after addition of chemokine. The cells were perfused at an estimated shear stress of 2.0 dynes cm^{-2} (flow rate of 0.78 ml min^{-1}). The entire period of perfusion was recorded on tape using a video recorder equipped with a time–date generator with a millisecond clock.

Calcium flux in monocytes. Elutriated human monocytes (10^7 ml^{-1}) were loaded with $1.0 \mu\text{M}$ Fura-2 AM (Molecular Probes) in the dark for 15 min at 37°C in DPBS containing 0.1% HSA, and intracellular calcium was measured as described²⁸. The data are presented as the relative ratio of fluorescence at 340 and 380 nm.

Chemotaxis assays. Chemotaxis assays were carried out in a 48-well microchemotaxis chamber (NeuroProbe) at 37°C for 90 min²⁹ on elutriated monocytes in HBSS buffer (MediaTech) supplemented with 0.05% low endotoxin BSA (Sigma) at 5×10^6 cells ml^{-1} . Monocytes that migrated across the filter and adhered to the bottom side were stained with Diff-Quick (Baxter Scientific). The cells per three 400 $\times$ fields in duplicate wells were counted and the data expressed as the mean $\pm$ standard deviation.

Statistical analysis. Data are expressed as the mean $\pm$ standard deviation or standard error of the mean, as indicated. Statistical comparison of means was performed by two-tailed unpaired Student's *t*-test. The null hypothesis was considered to be rejected at $P < 0.05$.

Received 4 February; accepted 2 March 1999.

- Smith, J. D. *et al.* Decreased atherosclerosis in mice deficient in both macrophage colony-stimulating factor (op) and apolipoprotein E. *Proc. Natl Acad. Sci. USA* **92**, 8264–8268 (1995).
- Suzuki, H. *et al.* A role for macrophage scavenger receptors in atherosclerosis and susceptibility to infection. *Nature* **386**, 292–296 (1997).
- Fazio, S. *et al.* Increased atherosclerosis in mice reconstituted with apolipoprotein E null macrophages. *Proc. Natl Acad. Sci. USA* **94**, 4647–4652 (1997).
- Nelken, N. A., Coughlin, S. R., Gordon, D. & Wilcox, J. N. Monocyte chemoattractant protein-1 in human atherosclerotic plaques. *J. Clin. Invest.* **88**, 1121–1127 (1991).
- Koch, A. E. *et al.* Enhanced production of the chemotactic cytokines interleukin-8 and monocyte chemoattractant protein-1 in human abdominal aortic aneurysms. *Am. J. Pathol.* **142**, 1423–1431 (1993).
- Boisvert, W. A., Santiago, R., Curtiss, L. K. & Terkeltaub, R. A. A leukocyte homologue of the IL-8 receptor CXCR-2 mediates the accumulation of macrophages in atherosclerotic lesions of LDL receptor-deficient mice. *J. Clin. Invest.* **101**, 353–363 (1998).
- Boring, L., Goslin, J., Cleary, M. & Charo, I. Decreased lesion formation in CCR2 $^{-/-}$ mice reveals a role for chemokines in the initiation of atherosclerosis. *Nature* **394**, 894–897 (1998).
- Baggiolini, M., Dewald, B. & Moser, B. Interleukin-8 and related chemokines—CXC and CC chemokines. *Adv. Immunol.* **55**, 97–179 (1994).
- Gerszten, R. E. *et al.* Adhesion of monocytes to vascular cell adhesion molecule-1-transduced human endothelial cells: implications for atherogenesis. *Circ. Res.* **82**, 871–878 (1998).
- Berkhout, T. A. *et al.* Cloning, *in vitro* expression, and functional characterization of a novel human CC chemokine of the monocyte chemoattractant protein (MCP) family that binds and signals through the CC chemokine receptor 2B. *J. Biol. Chem.* **272**, 16404–16413 (1997).
- Springer, T. A. Traffic signals for lymphocyte recirculation and leukocyte emigration: the multistep paradigm. *Cell* **76**, 301–314 (1994).
- Weber, C., Alon, R., Moser, B. & Springer, T. A. Sequential regulation of alpha 4 beta 1 and alpha 5 beta 1 integrin avidity by CC chemokines in monocytes: implications for transendothelial chemotaxis. *J. Cell Biol.* **134**, 1063–1073 (1996).
- Woldemar-Carr, M., Alon, R. & Springer, T. A. The C-C chemokine MCP-1 differentially modulates the avidity of $\beta 1$ and $\beta 2$ integrins on T lymphocytes. *Immunity* **4**, 179–187 (1996).
- Yoshimura, T. *et al.* Purification of a human monocyte-derived neutrophil chemotactic factor that has peptide sequence similarity to other host defense cytokines. *Proc. Natl Acad. Sci. USA* **84**, 9233–9237 (1987).
- Schroder, J. M., Mrowietz, U., Morita, E. & Christophers, E. Purification and partial biochemical characterization of a human monocyte-derived, neutrophil-activating peptide that lacks interleukin 1 activity. *J. Immunol.* **139**, 3474–3483 (1987).
- Chuntharapai, A., Lee, J., Hebert, C. A. & Kim, K. J. Monoclonal antibodies detect different distribution patterns of IL-8 receptor A and IL-8 receptor B on human peripheral blood leukocytes. *J. Immunol.* **153**, 5682–5688 (1994).
- Premack, B. A. & Schall, T. J. Chemokine receptors: gateways to inflammation and infection. *Nat. Med.* **2**, 1174–1178 (1996).
- Jones, K. A. *et al.* GABA(B) receptors function as a heteromeric assembly of the subunits GABA(B)R1 and GABA(B)R2. *Nature* **396**, 674–679 (1998).
- Luscinskas, F. W. *et al.* Monocyte rolling, arrest and spreading on IL-4-activated vascular endothelium under flow is mediated via sequential action of L-selectin, beta 1-integrins, and beta 2-integrins. *J. Cell Biol.* **125**, 1417–1427 (1994).
- Gerszten, R. *et al.* Adhesion of memory lymphocytes to VCAM-1-transduced human vascular endothelial cells under simulated physiological flow conditions *in vitro*. *Circ. Res.* **79**, 1205–1215 (1996).
- Yoshida, M., Szent, B. E., Kiely, J.-M., Rosenzweig, A. & Gimbrone, M. A. Regulation of E-selectin phosphorylation during leukocyte adhesion: a novel outside-in signaling role for E-selectin. *J. Immunol.* **161**, 933–941 (1997).
- Graber, N. *et al.* T cells bind to cytokine-activated endothelial cells via a novel, inducible sialoglycoprotein and endothelial leukocyte adhesion molecule-1. *J. Immunol.* **145**, 819–830 (1990).
- Bevilacqua, M. P., Prober, J. S., Mendrick, D. L., Cotran, R. S. & Gimbrone, M. A. Identification of an inducible endothelial-leukocyte adhesion molecule. *Proc. Natl Acad. Sci. USA* **84**, 9238–9242 (1987).
- Luscinskas, F. W. *et al.* Cytokine-activated human endothelial monolayers support enhanced neutrophil transmigration via a mechanism involving both endothelial-leukocyte adhesion molecule-1 and intercellular adhesion molecule-1. *J. Immunol.* **146**, 1617–1625 (1991).
- Taichman, D. B. *et al.* Tumor cell surface alpha 4 beta 1 integrin mediates adhesion to vascular endothelium: demonstration of an interaction with the N-terminal domains of INCAM-110/VCAM-1. *Cell Regul.* **2**, 347–355 (1991).
- Luscinskas, F. W. *et al.* L- and P-selectins, but not CD49d (VLA4) integrins, mediate monocyte initial attachment to TNF-activated vascular endothelium *in vitro*. *J. Immunol.* **156**, 326–335 (1996).
- Shen, J., Luscinskas, F. W., Connolly, A., Dewey, C. F. J. & Gimbrone, M. A. Fluid shear stress modulates cytosolic free calcium in vascular endothelial cells. *Am. J. Physiol.* **262**, C384–390 (1992).
- Wheeler, M. E., Luscinskas, F. W., Bevilacqua, M. P. & Gimbrone, M. A. Cultured human endothelial cells stimulated with cytokines or endotoxin produce an inhibitor of leukocyte adhesion. *J. Clin. Invest.* **82**, 1211–1218 (1988).

29. Falk, W., Goodwin, R. H. Jr & Leonard, E. J. A 48-well micro chemotaxis assembly for rapid and accurate measurement of leukocyte migration. *J. Immunol. Meth.* **33**, 239–247 (1980).

Acknowledgements. We thank K. Case and W. Atkinson for preparation of human endothelial cell cultures; L. Li and Q. Qui for technical assistance; P. Kaltfen for help with manuscript preparation; M. Cybulsky for rabbit VCAM-1 cDNA and antibody Rb1/9; D. Dichek for help in construction of AdvCAM-1 and for AdRSV β -gal; Leukosite for antibody 2D7; I. Charo for reviewing the manuscript; Z. N. Demirjian and other members of the MGH Clinical Hematology Laboratory for their review of leukocyte morphology. We acknowledge support from the NIH (R.E.G., A.R., M.A.G., F.W.L., A.D.L.). A.R. is an Established Investigator of the American Heart Association.

Correspondence and requests for materials should be addressed to A.R. (e-mail: rosenzweig@helix.mgh.harvard.edu).

Tyro-3 family receptors are essential regulators of mammalian spermatogenesis

Qingxian Lu^{††}, Martin Gore^{†††}, Qing Zhang[§], Todd Camenisch^{||}, Sharon Boast[§], Franca Casagrande[†], Cary Lai[#], Michael K. Skinner^{*}, Rüdiger Klein[†], Glenn K. Matsushima^{||}, H. Shelton Earp^{**}, Stephen P. Goff[§] & Greg Lemke^{*}

^{*} Molecular Neurobiology Laboratory, Salk Institute for Biological Studies, La Jolla, California 92037, USA

[§] Howard Hughes Medical Institute and Department of Biochemistry & Molecular Biophysics, Columbia University College of Physicians and Surgeons, New York, New York 10032, USA

^{||} Neuroscience Center and Department of Microbiology & Immunology, and ^{**} Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, North Carolina 27599, USA

[†] Developmental Biology Programme, European Molecular Biology Laboratory, 69117 Heidelberg, Germany

[#] Department of Neuropharmacology, Scripps Research Institute, La Jolla, California 92037, USA

^{*} Center for Reproductive Biology, Department of Genetics and Cell Biology, Washington State University, Pullman, Washington 99164, USA

^{††} These authors contributed equally to this work.

We have generated and analysed null mutations in the mouse genes encoding three structurally related receptors with tyrosine kinase activity: Tyro 3, Axl, and Mer^{1–4}. Mice lacking any single receptor, or any combination of two receptors, are viable and fertile, but male animals that lack all three receptors produce no mature sperm, owing to the progressive death of differentiating germ cells. This degenerative phenotype appears to result from a failure of the tropic support that is normally provided by Sertoli cells of the seminiferous tubules, whose function depends on testosterone and additional factors produced by Leydig cells^{5–7}. Tyro 3, Axl and Mer are all normally expressed by Sertoli cells during postnatal development, whereas their ligands, Gas6 and protein S, are produced by Leydig cells before sexual maturity, and by both Leydig and Sertoli cells thereafter. Here we show that the concerted activation of Tyro 3, Axl and Mer in Sertoli cells is critical to the role that these cells play as nurturers of developing germ cells. Additional observations indicate that these receptors may also be essential for the tropic maintenance of diverse cell types in the mature nervous, immune and reproductive systems.

The receptor protein-tyrosine kinases (PTKs) of the mammalian Tyro 3 family⁸ include Tyro 3 (also named Rse, Sky, Brt, Tif, Dtk, Etk-2)^{2,9,10}, Axl (also named Ark, Ufo, Tyro 7)^{3,11,12} and Mer (also named Eyk, Nyk, Tyro 12)^{4,13}. These three receptors are widely expressed in adult tissues, but their function is unknown. They share a distinctive structure, with extracellular regions composed of two immunoglobulin-related domains linked to two fibronectin type-III repeats, and cytoplasmic regions that contain an intrinsic PTK domain. Tyro 3, Axl and Mer are present in variable amounts in

^{††} Present address: Arena Pharmaceuticals, San Diego, California 92121, USA (M.G.).

neural, lymphoid, vascular and reproductive tissues, and in primary and tumour cell lines derived from these sources^{1,3,10–13}. The kinase activity of each of the receptors is activated by Gas6 (refs 14–17), a ligand that exhibits sequence relatedness to a steroid hormone transport protein designated the sex-hormone-binding globulin¹⁸. Tyro 3 can also bind and be activated by protein S, an anticoagulant in the blood coagulation cascade whose structure is closely related to that of Gas6 (ref. 14), although the extent to which protein S functions as a Tyro 3 ligand *in vivo* is debated¹⁵.

We investigated the function of Tyro 3, Axl and Mer by inactivating their genes in the mouse¹⁹ (Fig. 1). Mice homozygous for any single gene mutation were viable and fertile, and displayed no gross anatomical defects. Although activity-induced seizures have been observed in Tyro 3 single mutants more than seven months old (Q.L., M.G. and G.L., data not shown), and there is a two-fold increase in spleen size with an accompanying deficit in monocyte response to endotoxins in Mer homozygotes²⁰, all single mutants were obtained at the expected Mendelian frequency at weaning, and

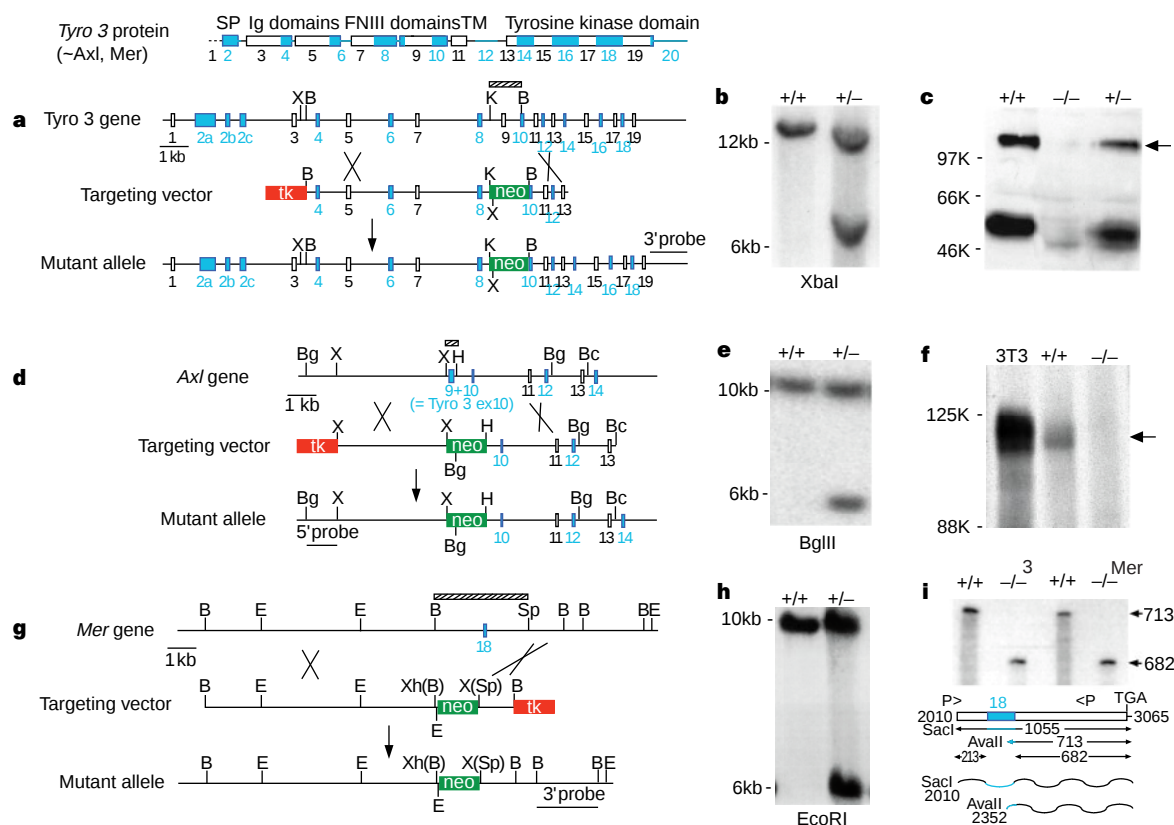


Figure 1 Inactivation of the mouse *Tyro 3*, *Axl* and *Mer* genes. The shared domain structure of these receptors is shown schematically (top); numbers denote exons of the mouse *Tyro 3* gene²⁷. **a**, Structure of the mouse *Tyro 3* gene, the targeting vector for inactivation, and the recombined allele. Signal peptide (SP) exons are numbered according to ref. 28, remaining exons are positioned according to ref. 27, with designations one integer lower after exon 3 (for example, exon 7 here is exon 8 of ref. 27). The 3' probe used for detection of homologous recombination of the right arm of the targeting vector is indicated. Ig, immunoglobulin-like; FNIII, fibronectin type III-like; TM, transmembrane domain; neo, G418 positive selection cassette; tk, thymidine kinase negative selection cassette; B, *Bam*HI; K, *Kpn*I; X, *Xba*I. **b**, Southern blot of *Xba*I-digested DNA from ES cells homozygous (+/+) or heterozygous (+/-) for the normal *Tyro 3* gene, hybridized with the 3' probe indicated in **a**. **c**, Western blot of brain protein lysates from wild-type (+/+), homozygous mutant (-/-), or heterozygous mutant mice (+/-), probed with a Tyro 3-specific antibody². Tyro 3 is arrowed. **d**, Structure of the *Axl* gene, the targeting vector for inactivation, and the recombined allele. Only the indicated exons have been mapped, although the targeting vector is likely to contain additional exons within its flanking arms, based on the structure of the human *Axl* gene²⁹. The 5' probe used for detection of homologous recombination of the left arm of the targeting vector is indicated. Bg, *Bgl*II; Bc, *Bcl*I; H, *Hind*III. **e**, Southern blot of *Bgl*II-digested DNA from ES cells homozygous (+/+) and heterozygous (+/-) for the normal *Axl* gene, hybridized with the 5' probe indicated in **d**. **f**, Western blot of total brain protein lysates prepared from wild-type (+/+) or homozygous mutant (-/-) mice, immunoprecipitated with an Axl-specific antibody³⁰ and then probed with the same antibody. NIH3T3 cells, which express high levels of the receptor³⁰, were similarly analysed as a positive control. Axl is indicated by an arrow. **g**, Restriction map of the mouse *Mer* gene and structure of

the targeting vector for the gene inactivation and of the recombined allele. The *Mer* exon that corresponds to exon 18 of the mouse *Tyro 3* gene (exon 19 of ref. 27) is contained within the indicated 3.1-kb *Bam*HI-*Spe*I (B-Sp) restriction fragment, but its precise location is unknown. Other *Mer* exons have not been mapped. The 53 and 1/3 amino acids of exon 18 extend from Met 725 in kinase sub-domain VI to Val 778 in sub-domain X. The 3' probe used for detection of homologous recombination of the right arm of the targeting vector is indicated. E, *Eco*RI; Xh, *Xho*I. **h**, Southern blot of *Eco*RI-digested tail DNA from mice homozygous (+/+) and heterozygous (+/-) for the *Mer* gene, hybridized with the 3' probe indicated in **g**. **i**, RNase protection analysis (RPA) of RNAs isolated from adult wild-type (+/+), *Mer* single mutant (-/-^{Mer}), and triple mutant (-/-³) mouse testes. Two radio-labelled RPA probes (wavy lines) were used: one runs from a *Sac*I site at nucleotide 2010 of the mouse *Mer* cDNA⁴ to nucleotide 3,065 (just downstream of the *Mer* TGA stop codon), and the second from an *Av* site at nucleotide 2,352 to nucleotide 3,065. RPA results using the *Av* probe are shown. This probe is fully protected from RNase digestion following hybridization to wild-type RNA (713-bp band), but is trimmed to a 682-bp fragment following hybridization to either *Mer* single or triple mutant RNA (arrowed lines). Similar RPAs with the *Sac*I probe (gels not shown) yielded the same 682-bp protected fragment upon hybridization to *Mer* mutant RNA, and also a protected fragment of 213 bp (arrowed). To confirm that these corresponded to deletion of the indicated *Mer* exon (blue), we performed RT-PCR on *Mer* mutant RNA using the indicated PCR primers (P> and <P). Sequencing of the amplified product demonstrated that only this exon (equivalent to *Tyro 3* exon 18 in **a**, or exon 19 of ref. 27) is deleted in the mutant. This 160-nucleotide deletion removes residues in kinase subdomains VII-IX, and portions of VI and X, which are essential to the activity of all known tyrosine kinases.

all lived for a long time in the laboratory. These results indicate that no single receptor of the Tyro 3 family is essential during embryonic development.

Because individual members of receptor PTK families often form heterodimers with one another to yield hybrid receptors with overlapping binding specificities, and co-expression of Tyro 3 and Axl or Mer has been observed for several cell types, we generated Tyro 3/Axl and Tyro 3/Mer double mutants. These mice were also obtained in the expected Mendelian ratios (we found that the *Tyro 3* and *Mer* genes are linked in the mouse genome) and were also viable and superficially healthy (but see below). Many double mutants, both male and female, were fertile as young adults, which allowed us to generate individuals that were mutated for all three receptor genes. Remarkably, even these triple knockouts were viable, and several have been maintained in our colony for over one year.

Their viability notwithstanding, the triple mutants displayed multiple major organ defects, neurological abnormalities and physiological deficits. Altered histology, increased apoptosis and cellular degeneration, for example, were prominent in a variety of adult tissues, including the hippocampus, cerebellum and neocortex of the brain, the epithelium of the prostate, the parenchyma of the liver, and the walls of blood vessels. In addition, adult triple mutants were blind, owing to the postnatal degeneration of rods and cones in the retina (Q.L. and G.L., data not shown). The spleens of adult triple mutants were populated by apoptotic cells but were also grossly enlarged, with an average weight of 468 ± 79 mg ($n = 4$) compared to 93 ± 8 mg ($n = 3$) for wild type, consistent with aberrant homeostasis of the lymphoid populations in which expression of the three receptors is prominent^{3,4,21}, and with the demonstration that protein S is interleukin-4-inducible in T lymphocytes and is an inhibitor of cultured B- and T-cell proliferation²¹. These observations demonstrate that concerted activation of Tyro 3 family receptors is essential for the trophic maintenance and homeostatic balance of a wide variety of cell types in mature mammalian tissues.

Among the most severe phenotypes was that seen in the adult male gonads. Although the ovaries of triple-knockout females were histologically abnormal, with cell death evident in the granulosa cells associated with growing ovarian follicles (Q.L. and G.L., data not shown), several of these females were fertile. In contrast, triple

knockout males never sired offspring. The testes of adult triple mutants were one-third of the size of those of wild-type males, with an average weight of 35 ± 8 mg ($n = 6$), compared to 100 ± 5 mg ($n = 5$) for wild type; triple-mutant epididymi were invariably devoid of sperm. Spermatogenesis was clearly disrupted in the triple mutants, and the cellular organization of the seminiferous tubules was grossly perturbed (Fig. 2). These effects were first detectable only at three weeks after birth, approximately one week before the onset of sexual maturity, and embryonic and early neonatal development of the testes, including the generation of a full complement of normal testicular cell types, was not obviously affected. In young (4–8 weeks) adults, a reduced number of round spermatids was first noticeable at seminiferous tubule stages VII and VIII (refs 7, 22); many fewer spermatocytes at the pachytene stage of meiosis were also seen in these young males, and there were very few cells with the morphological features of mature sperm (Fig. 2a, b). The release of the few mature sperm that were produced was delayed beyond tubule stage VIII, the normal stage of release²². These deficiencies became progressively more severe as triple-mutant males aged (Fig. 2c, d); in 6-month old mice, we frequently observed tubules depleted of nearly all germ cells (Fig. 2e–h). Apoptotic cell death, detected by TUNEL staining, was markedly elevated in the young triple-mutant tubules (Fig. 2i, j), and was prominent in spermatids at tubule stages VII–VIII and in pachytene spermatocytes (the same points in spermatid and spermatocyte differentiation at which cell death is first observed following hypophysectomy and testosterone depletion^{7,23,24}). Although apoptosis eventually depleted the testes of most germ cells, the proliferation of male stem cells (spermatogonia) continued for a time in young triple knockouts (Fig. 2k, l). Mutants carrying fewer receptor-gene mutations were also compromised with respect to spermatogenesis, although less severely. *Tyro 3*^{−/−}*Axl*^{+/+}*Mer*^{−/−} and *Tyro 3*^{−/−}*Axl*^{+/+}*Me*^{−/−} males had reduced testicular masses relative to wild type (52 and 43 mg, respectively), and only ~20% of *Tyro 3*^{−/−}*Axl*^{+/+}*Mer*^{−/−} males in our colony have sired offspring. These less-compromised mice also showed disordered spermatogenesis and reduced epididymal numbers of mature sperm. Less severe effects on fertility and testicular size, mass and histology were observed in *Tyro 3*^{−/−}*Axl*^{−/−}*Mer*^{+/+} males, and all single-mutant males appeared normal.

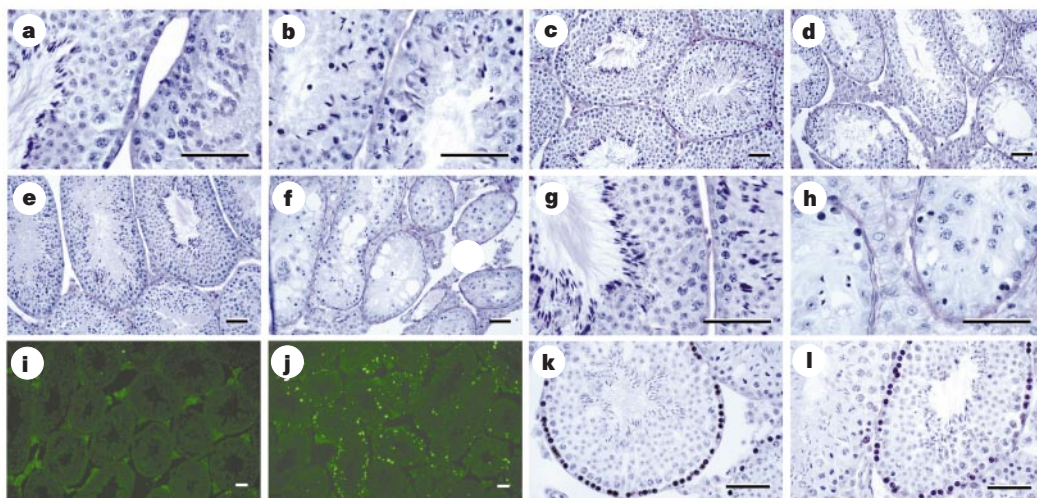


Figure 2 Aborted spermatogenesis and germ-cell death in seminiferous tubules of *Tyro 3*^{−/−}*Axl*^{−/−}*Mer*^{−/−} mice. **a, b**, Wild-type and triple-mutant seminiferous tubules, respectively, at 5 weeks after birth. The wild-type tubule section to the left of **a** is at late stage VII and contains many mature sperm. No mature sperm are evident in either of the two tubule sections in **b**, which contain stunted immature spermatids with improperly oriented acrosomes, and dying cells. **c, d**, Wild-type and triple-mutant tubules, respectively, at 18 weeks. Total cellularity of the triple-mutant tubules (**d**) is by this age substantially reduced. **e, f**, Wild-type and triple-

mutant tubules, respectively, at 6 months. Some tubule sections (such as the central tubule in **f**) are entirely depleted of germ cells in the triple mutants. **g, h**, Wild-type and triple-mutant tubules, respectively, at 6 months. **i, j**, TUNEL staining for apoptotic cell-death in sections of wild-type and triple-mutant tubules, respectively, at 9 weeks; **i** contains a single TUNEL⁺ cell; **j** contains 172 TUNEL⁺ cells. **k, l**, Spermatogonial stem cell division, as measured by a 2-h pulse of BrdU, in wild-type and triple-mutant seminiferous tubules, respectively, at 5 weeks. All scale bars, 50 μm.

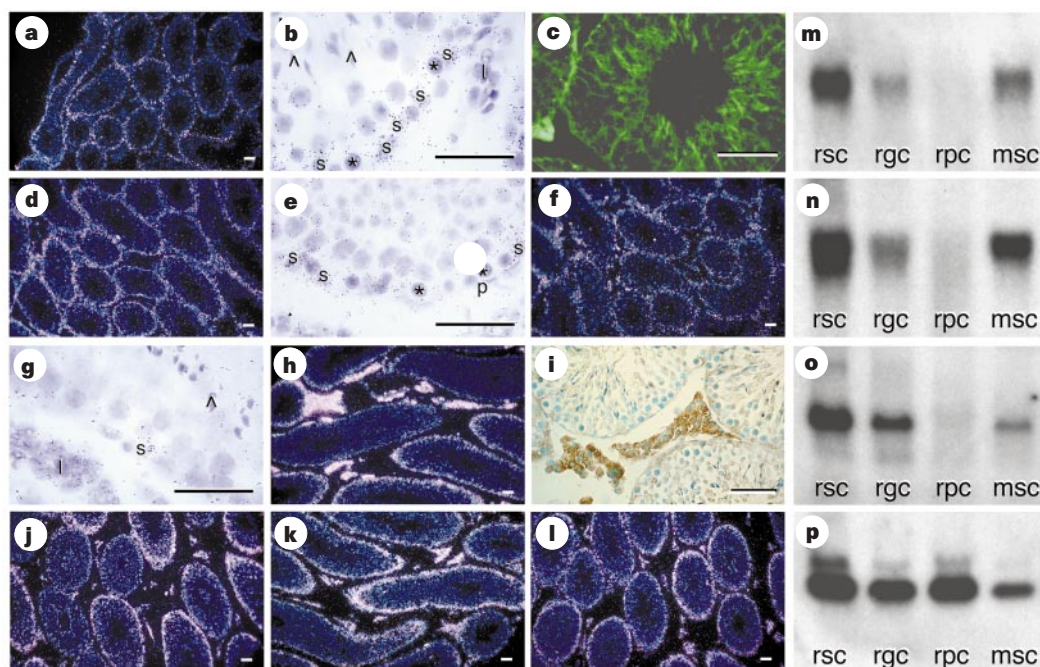


Figure 3 Tyro 3 family receptors and their ligands in the testes. **a**, *In situ* hybridization for Tyro 3 mRNA, expressed by cells at the periphery of all seminiferous tubule sections at 5 weeks. **b**, High-power, bright-field image of Tyro 3 mRNA (dark silver grains). Grains are located over lightly staining, irregularly shaped nuclei at the wall of the tubule, which are features of Sertoli-cell nuclei (s)²⁵, and not over the dark round nuclei of spermatogonia (asterisks), the Leydig cell nuclei outside the tubule (l), or spermatocyte or spermatid nuclei (A). **c**, Immunolocalization of Tyro 3 protein² (green) on most Sertoli cell membrane surfaces. **d**, *In situ* hybridization for Axl mRNA, similar to the Tyro 3 profile in **a**. **e**, High-power, bright-field image of Axl mRNA localization, similar to the Tyro 3 profile in **b**. Note the absence of signal in peritubular cells (p). **f**, *In situ* hybridization for Mer mRNA, similar to the Tyro 3 profile in **a**. **g**, High-power, bright-field image of Mer mRNA localization. Note low signal in Leydig cells, not seen for Tyro 3 or Axl mRNA. **h**, *In situ* hybridization for Gas6 mRNA, highly expressed by Leydig

cells, at 5 weeks. At this age, expression of Gas6 also first becomes detectable in a subset of Sertoli cells. **i**, Immunolocalization of Gas6 protein expression in Leydig cells at 4 weeks. **j**, *In situ* hybridization for Gas6 mRNA at 18 weeks. By this age, expression in Leydig and Sertoli cells is roughly equivalent, with a tubule-stage-specific regulation evident in the latter. (Sertoli cell expression is lowest at stages VII/VIII, the inverse of GATA-1 regulation.) **k**, *In situ* hybridization for protein S mRNA at 5 weeks, similar to the Gas6 profile in **h**. Although protein S mRNA is prominent in Leydig cells, it is also detected in Sertoli cells by this age. **l**, *In situ* hybridization for protein S mRNA at 18 weeks, similar to the Gas6 profile in **j**. **m-p**, Northern blots of total RNAs isolated from rat Sertoli cells (rsc), rat germ cells (rgc), rat peritubular cells (rpc), and mouse Sertoli cells (msc). Blots were hybridized with probes for Tyro 3 (**m**), Axl (**n**), SHBG, a marker for Sertoli cells (**o**), and cyclophilin as a loading/RNA control (**p**). All scale bars, 50 μ m.

Although spermatids and spermatocytes are the cells most obviously affected by the loss of Tyro 3, Axl and Mer, these cells do not in fact express the three receptors (Fig. 3). Before and after sexual maturity, each of the receptor messenger RNAs was localized to the periphery of the seminiferous tubules (Fig. 3a, d, f). Of the three cell types with nuclei at this position (Sertoli cells, spermatogonia and peritubular cells) only Sertoli cells clearly expressed Tyro 3, Axl and Mer mRNAs. This assignment was based on: (1) the absence of an *in situ* hybridization signal in the thin peritubular cells that surround the tubule (Fig. 3b, e, g); (2) the patchy circumferential distribution of receptor *in situ* signals and their association with the lightly staining, irregularly shaped nuclei characteristic of Sertoli cells²⁵, as opposed to the darkly staining, round nuclei of spermatogonia (Fig. 3a–g); (3) the presence of receptor mRNAs in cells that express sex-hormone-binding globulin, a Sertoli cell marker⁵ (Fig. 3m–o); and (4) the distribution of northern blot signals in RNAs isolated from enriched populations of testicular cells (Fig. 3m–o). In addition, antibodies raised against Tyro 3 (ref. 2) specifically stained Sertoli cells, which are large secretory cells whose cytoplasm extends across the full radial extent of the tubule²⁵ (Fig. 3c). Together, these data demonstrate that Tyro 3, Axl and Mer are Sertoli-cell products, although low-level expression by spermatogonia cannot be definitively excluded. Sertoli-cell expression of the receptor mRNAs was observed throughout postnatal development, with a modest peak at five weeks after birth (Fig. 3a, d, f). At all times, the amount of mRNA was greatest for Tyro 3 and Axl, and less for Mer. Unlike Tyro 3 and Axl mRNAs, a low level of

Mer mRNA was also detected in the testosterone-producing Leydig cells that lie outside the tubules (Fig. 3g).

In contrast to the receptors, their ligands were each expressed during the first three postnatal weeks by Leydig cells (Fig. 3h–l). There was a pronounced spike in Leydig cell Gas6 at the onset of sexual maturity (Fig. 3h, i); this expression declined and stabilized in mature adults (Fig. 3j). Co-incident with sexual maturation, Gas6 mRNA and protein also appeared in Sertoli cells (Fig. 3h). This Sertoli-cell signal was first detected at stage IX during the first wave of spermatogenesis to pass through the tubule, and became increasingly prominent with postnatal age, such that in mature testes roughly comparable levels of Gas6 were detected in Leydig and Sertoli cells (Fig. 3j). Sertoli-cell expression of Gas6 in the mature testes exhibited a distinctive rising-and-falling regulation as a function of tubule stage; preliminary analysis indicates that expression is highest at stages III–IV and is almost undetectable by stage VIII (Fig. 3j). This periodic cycling with tubule stage is a property exhibited by many other Sertoli-cell gene products, including the transcription factor GATA-1 (ref. 26) (see below). Protein S shows a similar development profile (Fig. 3k, l), although its overall expression was less than that of Gas6, and Sertoli cells became protein-S-positive about one week before they became Gas6-positive. Protein S levels also fluctuated with tubule stage, although the amplitude of this fluctuation was smaller than for Gas6. Thus, ligand–receptor signalling for the Tyro 3 family in the testes switches from a paracrine to an autocrine/paracrine mode at the onset of sexual maturity.

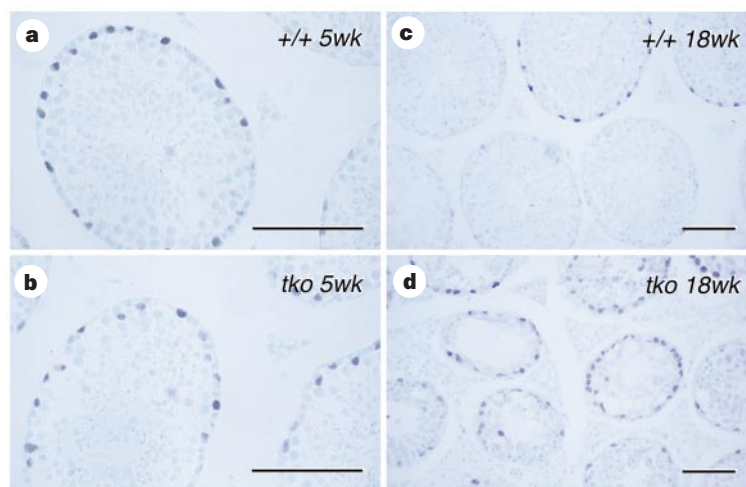


Figure 4 Expression of GATA-1 in wild-type (+/+) and triple-mutant (tko) testes. **a**, **b**, GATA-1 immunoreactive (dark blue) nuclei in sections of wild-type and triple-mutant seminiferous tubules, respectively, at 5 weeks. GATA-1⁺ Sertoli cells are present in the triple mutant at this age. **c**, **d**, GATA-1 immunoreactivity at 18 weeks.

These observations indicated that germ-cell death in the triple mutants might result from the death of Sertoli cells. This is not the case: we readily detected GATA-1⁺ Sertoli cells within the tubules of young adult triple knockouts (Fig. 4a, b), and the number of GATA-1⁺ cells significantly increased in more mature triple mutants relative to wild-type (Fig. 4c, d). Sertoli-cell expression of GATA-1 also fluctuates with tubule stage in the mature testes, with a peak at tubule stages VII/VIII (Fig. 3c)²⁶, and germ cells inhibit GATA-1 expression at other tubule stages²⁶. We found that, in the germ-cell-depleted tubules of older mutants, most Sertoli cells express GATA-1 independently of tubule stage (Fig. 4d).

Sertoli cells provide essential physical and trophic support for developing spermatogonia, spermatocytes and spermatids. This tropic support in turn depends on a network of reciprocal signalling interactions between Leydig cells, Sertoli cells, peritubular cells and germ cells within the testes⁵, and on upstream hormonal regulation from the pituitary. Although intratesticular interactions are well understood at a cellular level, the underlying signalling molecules have, with the exception of testosterone itself, remained elusive. Our results now identify the Tyro 3 family receptors and their ligands as essential regulators of Sertoli-cell function. The fact that the *Tyro 3*, *Axl* and *Mer* genes must all be disrupted for maximal germ-cell degeneration to occur indicates that these receptors act additively or in combination; their combined activation in Sertoli cells may be required for the production of the germ-cell trophins that these cells must produce. Although the spermatotoxic effects seen in the triple knockouts are in many respects similar to those following hypophysectomy or testosterone deprivation^{23,24}, other behavioural and physiological effects of these manipulations are not evident in the mutants: these mice exhibit normal sexual behaviour, have histologically intact seminal vesicles, and have circulating testosterone levels in the normal range – 75, 85 and 185 ng dl⁻¹ for three triple knockout males at 3 months compared to 48, 105 and 110 ng dl⁻¹ for three wild-type males of the same age.

As (1) Gas 6 and protein S are related in sequence and predicted structure to sex-hormone-binding globulin; (2) the primary cellular target of testosterone in the seminiferous tubule is the Sertoli cell; (3) Leydig cells produce both testosterone and Gas6/protein S; and (4) abrogation of signalling through Tyro 3 family receptors leads to germ-cell effects similar to those of testosterone depletion, we suggest that the actual ligands for Tyro 3 family receptors may be hybrid molecules of Gas6 or protein S bound to testosterone or (in other cells) other steroids. Given the large and controversial

GATA-1 expression persists in tubule sections of the triple mutant (**d**), and in contrast to wild-type (**c**), this expression is evident in nearly all tubules, independent of tubule stage. All scale bars, 100 μm.

literature on cell-surface, non-genomic effects of steroids, we are currently investigating this possibility. □

Methods

Gene inactivation. Each of the receptor genes was disrupted using a G418 (neomycin)-resistance cassette, and the inactivated genes were used to replace their normal counterparts by homologous recombination in mouse embryonic stem (ES) cells (Fig. 1). A *KpnI*–*Bam*HI (K–B) fragment containing exon 9 and a portion of exon 10 of the *Tyro 3* gene (striped bar in Fig. 1a) was replaced with the resistance cassette (neo), used for positive selection in ES cells¹⁹. A thymidine-kinase cassette (tk) was used to select against non-homologous recombinants. An *XbaI*–*Hind*III fragment containing exon 9 of the *Axl* gene (striped bar in Fig. 1d), and a 3.1-kb *Bam*HI–*Spe*I fragment of the *Mer* gene (striped bar in Fig. 1g) were both similarly replaced, and tk was in each case used for negative selection. ES cells heterozygous for the inactivated genes were injected into wild-type mouse blastocysts to generate chimaeras, which were then bred and intercrossed in subsequent generations to yield mice homozygous for each of the mutations, according to standard protocols¹⁹. Loss of functional Tyro 3, Axl or Mer in these mice was confirmed by western and/or RNA analysis (Fig. 1c, f, i; and data not shown). The primers used for the RT-PCR experiments of Fig. 1i were from nucleotides 1,963–1,984 (>, forward) and 2,736–2,757 (<, reverse) of the mouse Mer cDNA (ref. 4).

Expression studies. Gas6 and GATA-1 antibodies were purchased from Santa Cruz Biologicals. *In situ* hybridization, western blots, northern blots, and immunohistochemistry were performed as described^{2,8,14}. ³²P-riboprobes were used for *in situ* hybridizations, antibody binding to westerns was detected by ECL chemiluminescence, and northern blots were hybridized with ³²P-radiolabelled DNA probes. Seminiferous tubule-cell populations were separated by standard protocols^{5–7}. Isolation methods result in highly purified populations of Sertoli and peritubular cells, but germ-cell preparations contain some Sertoli cells, as shown by the distribution of sex-hormone-binding globulin signal (Fig. 3p). Exposure times for *in situ* hybridizations of Fig. 3: j, 12 days; z, b, d, e, h, k, l, 15 days; f, g, 21 days.

Received 19 November 1998; accepted 2 March 1999.

- Crosier, K. E. & Crosier, P. S. New insights into the control of cell growth; the role of the Axl family. *Pathology* **29**, 131–135 (1997).
- Lai, C., Gore, M. & Lemke, G. Structure, expression, and activity of Tyro 3, a neural adhesion-related receptor tyrosine kinase. *Oncogene* **9**, 2567–2578 (1994).
- O'Bryan, J. P. et al. Axl, a transforming gene isolated from primary human myeloid leukemia cells, encodes a novel receptor tyrosine kinase. *Mol. Cell. Biol.* **11**, 5016–5031 (1991).
- Graham, D. K., Dawson, T. L., Mullaney, D. L., Snodgrass, H. R. & Earp, H. S. Cloning and mRNA expression analysis of a novel human protooncogene, *c-mer*. *Cell Growth Differ.* **5**, 647–657 (1994).
- Skinner, M. K. Cell-cell interactions in the testis. *Endocrine Rev.* **12**, 45–77 (1991).
- Skinner, M. K. in *The Sertoli Cell* (eds Russell, L. D. & Griswold, M. D.) 237–247 (Cache River, Clearwater, Florida, 1993).

7. Griswold, M. D. Interactions between germ cells and Sertoli cells in the testis. *Biol. Reprod.* **52**, 211–216 (1995).
8. Lai, C. & Lemke, G. An extended family of protein-tyrosine kinase genes differentially expressed in the vertebrate nervous system. *Neuron* **6**, 691–704 (1991).
9. Mark, M. R. *et al.* RSE, a novel receptor-type tyrosine kinase with homology to Axl/Ufo, is expressed at high levels in the brain. *J. Biol. Chem.* **269**, 10720–10728 (1994).
10. Taylor, I. C., Roy, S., Yaswen, P., Stampfer, M. R. & Varmus, H. E. Mouse mammary tumors express elevated levels of RNA encoding the murine homology of SKY, a putative receptor tyrosine kinase. *J. Biol. Chem.* **270**, 6872–6880 (1995).
11. Rescigno, J., Mansukhani, A. & Basilico, C. A putative receptor tyrosine kinase with unique structural topology. *Oncogene* **6**, 1909–1913 (1991).
12. Janssen, J. W. *et al.* A novel putative tyrosine kinase receptor with oncogenic potential. *Oncogene* **6**, 2113–2120 (1991).
13. Jia, R. & Hanafusa, H. The proto-oncogene of v-ryk (v-ryk) is a novel receptor-type protein tyrosine kinase with extracellular Ig/GN-III domains. *J. Biol. Chem.* **269**, 1839–1844 (1994).
14. Stitt, T. N. *et al.* The anticoagulation factor protein S and its relative, Gas6, are ligands for the Tyro 3/Axl family of receptor tyrosine kinases. *Cell* **80**, 661–670 (1995).
15. Godowski, P. J. *et al.* Reevaluation of the roles of protein S and Gas6 as ligands for the receptor tyrosine kinase Rse/Tyro 3. *Cell* **82**, 355–358 (1995).
16. Chen, J., Carey, K. & Godowski, P. J. Identification of Gas6 as a ligand for Mer, a neural cell adhesion molecule related receptor tyrosine kinase implicated in cellular transformation. *Oncogene* **14**, 2033–2039 (1997).
17. Nagata, K. *et al.* Identification of the product of growth arrest-specific gene 6 as a common ligand for Axl, Sky, and Mer receptor tyrosine kinases. *J. Biol. Chem.* **271**, 30022–30027 (1996).
18. Joseph, D. R. Sequence and functional relationships between androgen-binding protein/sex hormone-binding globulin and its homologs protein S, Gas6, laminin, and agrin. *Steroids* **62**, 578–588 (1997).
19. Capecchi, M. R. Altering the genome by homologous recombination. *Science* **244**, 1288–1292 (1989).
20. Camenisch, T. D., Koller, B. H., Earp, H. S. & Matsushima, G. K. A novel receptor tyrosine kinase, Mer, inhibits TNF- α production and lipopolysaccharide-induced endotoxin shock. *J. Immunol.* **162**, 3498–3503 (1999).
21. Smiley, S. T., Stitt, T. N. & Grusby, M. J. Cross-linking of protein S bound to lymphocytes promotes aggregation and inhibits proliferation. *Cell. Immunol.* **181**, 120–126 (1997).
22. Leblond, C. P. & Clemont, Y. Definition of the stages of the cycle of the seminiferous epithelium in the rat. *Ann. N. Y. Acad. Sci.* **55**, 548–573 (1952).
23. Henriksen, K., Hakovirta, H. & Parvinen, M. Testosterone inhibits and induces apoptosis in rat seminiferous tubules in a stage-specific manner. *Endocrinology* **136**, 3285–3291 (1995).
24. Sun, Y.-T., Wreford, N. G., Robertson, D. M. & De Kretser, D. M. Quantitative cytological studies of spermatogenesis in intact and hypophysectomized rats: Identification of androgen-dependent stages. *Endocrinology* **127**, 1215–1223 (1991).
25. Wong, V. & Russell, L. D. Three-dimensional reconstruction of a rat stage V Sertoli cell: I. Methods, basic configuration, and dimensions. *Am. J. Anat.* **167**, 143–161 (1983).
26. Yomogida, K. *et al.* Developmental stage- and spermatogenic cycle-specific expression of transcription factor GATA-1 in mouse Sertoli cells. *Development* **120**, 1759–1766 (1994).
27. Lewis, P. A., Crosier, K. E., Wood, C. R. & Crosier, P. S. Analysis of the murine Dtk gene identifies conservation of genomic structure within a new receptor tyrosine kinase subfamily. *Genomics* **31**, 13–19 (1996).
28. Biesecker, L. G., Giannola, D. M. & Emerson, S. G. Identification of alternative exons, including a novel exon, in the tyrosine kinase receptor gene Etk2/tyro3 that explains differences in 5' cDNA sequences. *Oncogene* **10**, 2239–2242 (1995).
29. Schulz, A. S., Schleithoff, L., Faust, M., Bartram, C. R. & Janssen, J. W. G. The genomic structure of the human UFO receptor. *Oncogene* **8**, 509–513 (1993).
30. Bellosta, P., Zhang, Q., Goff, S. P. & Basilico, C. Signaling through the ARK tyrosine kinase receptor protects from apoptosis in the absence of growth stimulation. *Oncogene* **15**, 2387–2397 (1997).

Acknowledgements. This work was supported by grants from the NIH (to G.L., S.P.G., H.S.E., G.K.M.), and by a Markey fellowship to M.G., who was a student in the Neurosciences graduate program at the University of California, San Diego. S.P.G. is an Investigator of the Howard Hughes Medical Institute. We thank A. Prieto for advice and for Gas6 antibody, G. Yancopoulos for Gas6, protein S, and Axl cDNA probes, and D. Ortuño, P. Burrola, and D. Baynes for technical assistance.

Correspondence and requests for materials should be addressed to G.L. (e-mail: lemke@salk.edu).

Binding of double-strand breaks in DNA by human Rad52 protein

Eric Van Dyck*, Alicja Z. Stasiak†, Andrzej Stasiak† & Stephen C. West*

*Imperial Cancer Research Fund, Clare Hall Laboratories, South Mimms, Herts EN6 3LD, UK

†Laboratoire d'Analyse Ultrastructurale, Université de Lausanne, CH-1015 Lausanne-Dorigny, Switzerland

Double-strand breaks (DSBs) in DNA are caused by ionizing radiation. These chromosomal breaks can kill the cell unless repaired efficiently, and inefficient or inappropriate repair can lead to mutation, gene translocation and cancer¹. Two proteins that participate in the repair of DSBs are Rad52 and Ku: in lower eukaryotes such as yeast, DSBs are repaired by Rad52-dependent homologous recombination, whereas vertebrates repair DSBs

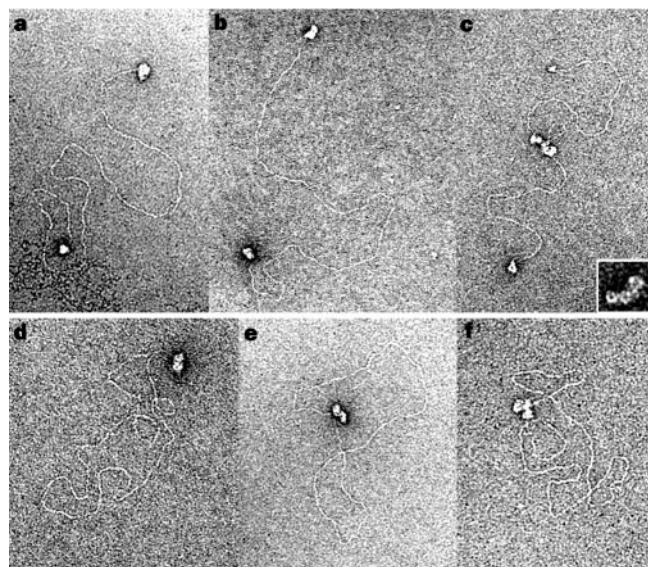


Figure 1 DNA end-binding and end-to-end interactions promoted by hRad52. Complexes formed between hRad52 and linear duplex DNA containing single-stranded tails were visualized by electron microscopy. **a, b** and **f**, 3' single-stranded (ss) DNA tails. **c, d** and **e**, 5' ssDNA tails. Inset in **c** shows a close-up view of an end-binding complex in which hRad52 rings are evident. Reactions (20 μ l) contained 10 μ M (**a, d**) or 4.7 μ M (**b, c, e, f**) tailed duplex DNA in 20 mM triethanolamine-HCl (pH 7.5). After 5 min at 37 °C, hRad52 was added to 0.16 μ M (**a, d**) or 0.08 μ M (**b, c, e, f**) and incubation was continued for 15 min.

primarily by Ku-dependent non-homologous end-joining². The contribution of homologous recombination to vertebrate DSB repair, however, is important^{3,4}. Biochemical studies indicate that Ku binds to DNA ends and facilitates end-joining⁵. Here we show that human Rad52, like Ku, binds directly to DSBs, protects them from exonuclease attack and facilitates end-to-end interactions. A model for repair is proposed in which either Ku or Rad52 binds the DSB. Ku directs DSBs into the non-homologous end-joining repair pathway, whereas Rad52 initiates repair by homologous recombination. Ku and Rad52, therefore, direct entry into alternative pathways for the repair of DNA breaks.

In lower eukaryotes, double-strand breaks in meiotic and mitotic cells are processed to form single-stranded tails, in reactions that involve the Mre11/Rad50/Xrs2 protein complex^{6,7}. Similar processing events are thought to occur in higher cells, where homologues of Mre11, Rad50 and Xrs2 (p95) exist². In yeast *rad52* mutants, exonuclease degradation is more extensive than in wild-type cells⁷, indicating that the Rad52 protein may be involved in the protection or processing of the initial DSB. We therefore tested whether Rad52 acts at the early stages of double-strand-break repair by direct interaction with the DSB or partially resected DSB. Complexes were formed between human Rad52 protein (hRad52) and linear duplex DNA, and were visualized by electron microscopy. Using substrates containing 5' or 3' single-stranded tails (approximately 300 nucleotides in length), we observed that hRad52 bound preferentially to the ends of the DNA (Fig. 1). Molecules in which both ends of the linear DNA were bound by hRad52 are shown in Fig. 1a–c.

We found that a large percentage of the linear DNA (~90%) present on each grid was held together by hRad52-mediated intermolecular interactions. Indeed, end-to-end associations between DNA molecules, mediated by hRad52, resulted in the formation of large DNA–protein networks (data not shown). However, away from the networks, isolated DNA molecules were observed in which Rad52 protein clearly provided an intramolecular bridge between two ends, resulting in their recircularization (Fig. 1d–f). When 46

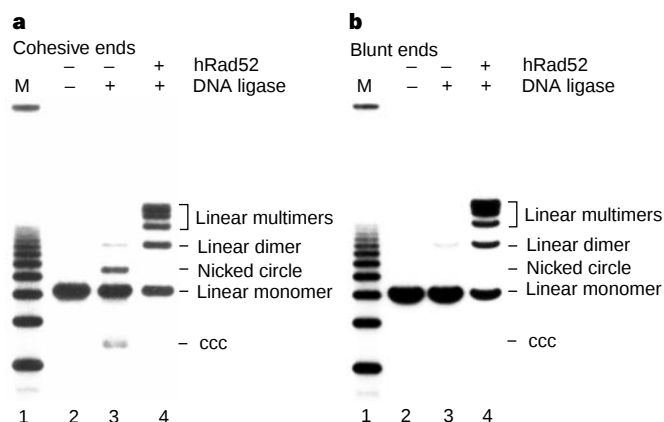


Figure 2 Stimulation of DNA ligation by hRad52. **a, b**, Effect of hRad52 on the ligation of *Eco*RI- or *Sma*I-linearized DNA, respectively. M, size markers (1-kb ladder). See Methods for experimental details.

well-spread individual 3'-tailed molecules were examined, we found that 34 were recircularized by hRad52, three contained hRad52 on both ends of a linear molecule, and nine contained hRad52 on one end only. Similarly, with 5'-tailed DNA, examination of 30 molecules revealed that 16 were recircularized, seven were bound at both ends and seven were bound at one end only. Although the self-association of hRad52 has been demonstrated by two-hybrid analysis⁸, our observations indicate a direct role for these protein-protein interactions in promoting both inter- and intramolecular DNA contacts. hRad52 exhibited a clear preference for binding DNA ends, and only rare molecules were observed in which hRad52 bound internally. In these instances, hRad52-hRad52 interactions occasionally led to 'loop' formation (Fig. 1c).

Electron microscopic studies of yeast⁹ and human¹⁰ Rad52 have revealed the formation of ring-shaped structures, both in solution and on DNA. Electron microscopic analyses of the end-binding complexes showed that they were generally amorphous in shape and ranged in size from approximately 15 to 60 nm. Within these complexes, hRad52 rings could occasionally be seen (Fig. 1c, inset).

Because hRad52 exhibits preferential binding to single-stranded DNA compared with double-stranded DNA^{10,11}, we next determined whether the single-stranded tails were required for DNA end-binding and end-to-end interactions promoted by hRad52. Protein-DNA complexes were prepared using blunt-ended linear duplex DNA. End-binding and end-to-end associations were again observed by electron microscopy (data not shown). With this substrate, however, the total number of molecules bound by hRad52 was reduced compared to that observed with the tailed substrates, and the complexes were generally smaller (~15–30 nm). These results indicate that hRad52 may have a higher affinity for (or greater stability on) tailed DNA than for blunt-ended linear duplex DNA.

Given that hRad52 causes end-to-end interactions, its effect on the efficiency of DNA ligation was investigated. Using linear DNA molecules containing cohesive (Fig. 2a) or blunt (Fig. 2b) ends, we found that preincubation of the DNA with hRad52 stimulated ligation catalysed by T4 DNA ligase. With the cohesive-ended DNA, most hRad52-facilitated events were intermolecular and approximately 65% of the DNA was converted into multimeric linear forms (Fig. 2a, lane 4). Analysis by digestion with restriction enzymes revealed that ligation involved both head-to-head and head-to-tail associations (data not shown). In the absence of hRad52, ligation resulted primarily in the formation of circular (covalently closed and nicked circular) species (Fig. 2a, lane 3). The stimulatory effect of hRad52 on DNA ligation was particularly

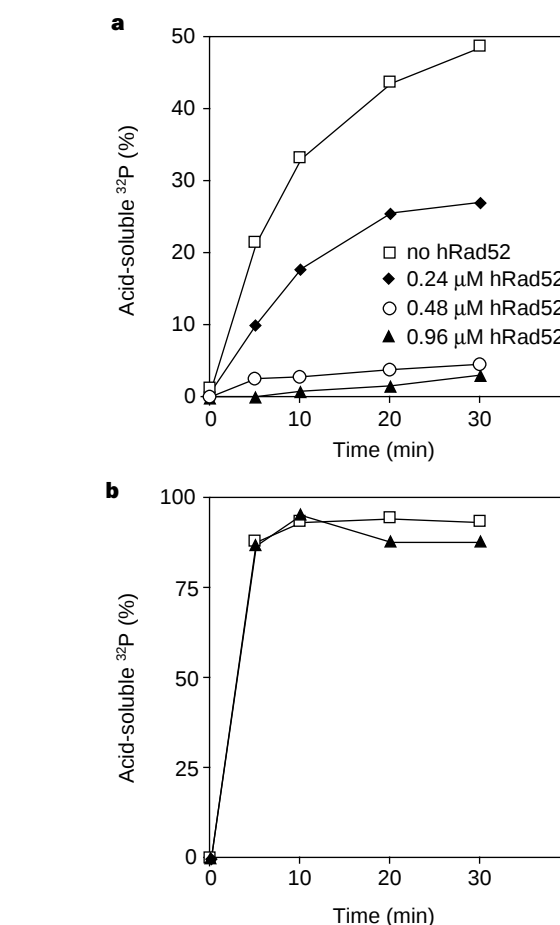


Figure 3 DNA-end protection by hRad52 protein. **a, b**, hRad52 was incubated with uniformly ³²P-labelled *Eco*RI-linearized DNA (30 μM), followed by the addition of either exonuclease III (**a**) or DNase I (**b**). Samples were taken at the indicated times and analysed for the release of acid-soluble ³²P counts as described in Methods. The final concentrations of hRad52 are indicated.

striking with blunt-ended DNA (Fig. 2b: compare lanes 3 and 4). Although our observation that end-ligation occurs in the presence of hRad52 may indicate that the ends remain freely accessible to further enzymatic processing, it is perhaps more likely that ligation results from Rad52-mediated network formation. Protein-mediated stimulation of end-ligation has also been observed with hRad51 protein¹².

Our discovery that hRad52 possesses DNA end-binding activity implies that the proteins should be capable of protecting linear DNA from exonuclease digestion. To test this, uniformly ³²P-labelled *Eco*RI-linearized duplex DNA (7 nM DNA ends) was incubated with hRad52 (0.24–0.96 μM) before the addition of *Escherichia coli* exonuclease III. Remarkably, hRad52 afforded almost complete protection from exonuclease attack (Fig. 3a). Indeed, less than 5% of the ³²P was released as acid-soluble counts at concentrations of hRad52 ≥ 0.48 μM. In contrast, using 0.96 μM hRad52, we failed to observe any protection from digestion by an endonuclease such as DNase I (Fig. 3b), indicating that the bulk of the DNA remained accessible. Similar results were obtained using blunt-ended DNA (data not shown).

We have shown that hRad52 proteins binds double-strand DNA breaks, protects them from exonuclease attack and promotes end-to-end interactions. These properties are remarkably similar to those exhibited by Ku⁵, leading us to propose a model for the repair of DSBs in human cells in which recognition of the DSB by

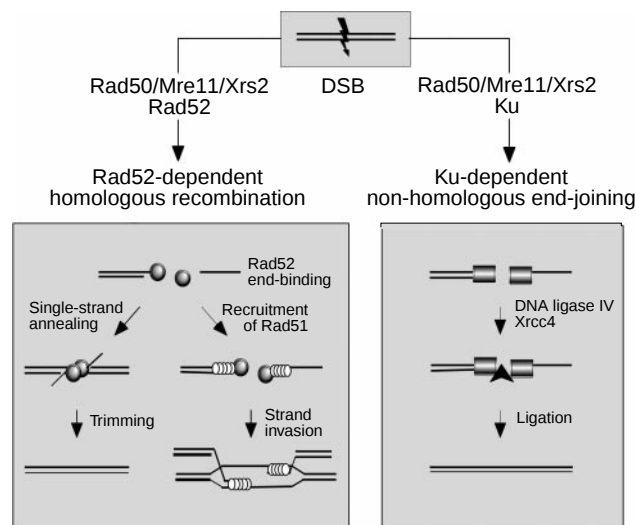


Figure 4 Model for the initiation of double-strand break repair. In this model, a DSB can be bound by hRad52, directing repair by homologous recombination (left), or by Ku, which initiates repair by non-homologous end-joining (right). Rad52-mediated repair can occur by single-strand annealing or by Rad51-mediated strand invasion of a homologous duplex DNA. Rad52 is indicated by shaded circles, Rad51 by white ellipses, and DNA-PK (that is, Ku/DNA-PK_{cs}) by shaded rectangles.

either hRad52 or Ku directs repair by Rad52-dependent homologous recombination or Ku-dependent non-homologous end-joining (Fig. 4).

That Rad52 and Ku provide alternative means to process DSBs is consistent with the phenotypes of *rad52/Ku* mutants. In higher organisms, the importance of Ku (in association with the DNA-activated protein kinase DNA-PK_{cs}) is illustrated by the X-ray-sensitive and V(D)J-recombination-deficient phenotype exhibited by homozygous mouse *Ku*^{-/-} knockouts^{13,14}. Because Ku is abundant¹⁵, defects in hRad52 alone fail to cause a severe phenotype, as indicated by the apparently normal radiation sensitivity found with *RAD52*^{-/-} knockout mouse¹⁶ and chicken¹⁷ cells. In yeast cells, however, where Ku-directed non-homologous end-joining represents only a minor pathway for DSB repair⁵, *rad52* mutants exhibit a severe recombination defect and are highly sensitive to ionizing radiation¹⁸.

Although vertebrate cells defective in Rad52 do not exhibit an extreme recombination/repair-defective phenotype, the contribution of homologous recombination to the repair of DSBs should not be underestimated. In hamster cells, endonuclease-generated DSBs in direct repeat sequences are repaired by homologous recombination³. Moreover, studies with the *RAD54*^{-/-}/*Ku70*^{-/-} double knockout from the chicken B-cell line DT40 show that, whereas non-homologous end-joining (NHEJ) is more important for repairing γ -radiation-induced DSBs during G1–early S phase, recombinational repair is preferentially used in late S–G2, when an intact sister chromatid is available⁴. The lack of a strong phenotype in the *RAD52*-knockout mouse may be due to a redundancy of Rad52 function, with some other function (for example, a Rad59 homologue³) taking over the end-binding role in the absence of Rad52.

In the model shown in Fig. 4, hRad52 binds DSBs induced in somatic cells by ionizing radiation or chemical damage. DSB-binding by hRad52 is equally applicable, however, at the initiation of recombination in germline cells at meiosis. Binding may occur before, during or after resection of the DSB by Mre11/Rad50/Xrs2, resulting in the formation of duplex molecules containing short single-stranded DNA tails bound by hRad52. The end-protection

property of hRad52 raises the possibility that Rad52 directly controls the extent of exonuclease resection, and offers an explanation for the more extensive exonuclease digestion observed in yeast *rad52* mutants compared with wild-type cells⁷.

Genetic studies in yeast have shown that *RAD52* is required for two homologous recombination pathways: (1) single-strand annealing which occurs between repeated DNA sequences, and (2) recombination promoted by the *RAD51*, *RAD54*, *RAD55* and *RAD57* gene products¹⁹. Consistent with a direct role in these processes, Rad52 anneals complementary single-strands^{9,20–22} and can stimulate *in vitro* recombination reactions promoted by Rad51 (refs 11, 23–25). Rad52, Rad51 and RP-A all co-localize during meiosis at sites of potential DSBs (ref. 26); also, Rad52 is required for the pre-assembly of recombination complexes containing Rad51 (ref. 26). Rad52 may help to target Rad51 to the site of the resected DSB, where it will initiate strand exchange with homologous duplex DNA. Assembly of the recombination complex at the site of bound Rad52 is also likely to involve Rad55, Rad57 and RP-A (refs 23–27).

The frequency of spontaneous allelic recombination in higher eukaryotic cells is extremely low²⁸, of the order of 10⁻⁸, so gene therapy by *in vivo* gene targeting will remain impractical unless a dramatic improvement in targeting efficiency can be achieved. The frequency of homologous recombination can be stimulated²⁹ and resistance to ionizing radiation can be increased³⁰ by overexpression of Rad52 in mammalian cells, consistent with the DSB-repair model presented here. Moreover, our model predicts that significant increases in the frequency of homologous gene targeting using linear DNA vectors may be achieved by downregulating or inactivating Ku in coordination with Rad52 overexpression. □

Methods

Proteins. Recombinant human Rad52 was purified from baculovirus-infected Sf9 cells as described previously, except that the Ni-NTA-agarose column was replaced with Talon¹⁰. hRad52 was stored and diluted in 20 mM Tris-HCl (pH 8.0), 100 mM KCl, 1 mM EDTA, 0.5 mM dithiothreitol, and 10% glycerol. T4 DNA ligase (NEB), exonuclease III (BRL), λ exonuclease (Pharmacia) and restriction enzymes were used as recommended by the suppliers.

DNA substrates. Uniformly ³²P-labelled plasmid pPB4.3 DNA (4,270 bp)¹¹ was prepared from cells grown in ³²P-orthophosphate. To produce linear duplex DNA with 5' or 3' single-stranded DNA tails, ³²P-labelled form-I plasmid DNA was cut with *Sma*I and digested with predetermined saturating amounts of either exonuclease III or λ exonuclease. Exonuclease treatment was controlled to produce 5' or 3' tails with average lengths of 350 (5' tail) or 280 nucleotides (3' tail). The extent of digestion was determined by measuring the release of acid-soluble ³²P counts after precipitation with perchloric acid. All DNA concentrations are expressed in moles of nucleotide residues.

Electron microscopy. Protein–DNA complexes were fixed by addition of glutaraldehyde to 0.2% followed by a 15 min incubation at 37 °C. Samples were diluted and washed in 5 mM magnesium acetate before uranyl acetate staining¹⁰. Complexes were visualized at magnifications of 20,500 $\times$ or 25,000 $\times$ using a Philips C100 electron microscope.

Stimulation of DNA ligation. hRad52 (0.32 μ M) was incubated for 15 min at 37 °C with ³²P-labelled linearized pPB4.3 DNA (10 μ M) in 20 mM triethanolamine-HCl (pH 7.5). To this mixture (10 μ l) was added 1 μ l of 10 $\times$ buffer (500 mM Tris-HCl (pH 7.5), 100 mM MgCl₂, 10 mM ATP, and 250 μ g ml⁻¹ bovine serum albumin) and either 40 (Fig. 2a) or 200 (Fig. 2b) units of T4 DNA ligase. Incubation was at room temperature for 2.5 hours. Reactions were stopped and deproteinized by addition of 2 μ l 'stop' buffer (2% SDS, 0.6 mg ml⁻¹ proteinase K), followed by 15 min incubation at 37 °C. Products were analysed by 0.7% agarose gel electrophoresis using TAE buffer containing 1 μ g ml⁻¹ ethidium bromide followed by autoradiography.

Exonuclease protection. Mixtures (120 μ l) containing hRad52 and ³²P-labelled linear duplex DNA in 20 mM triethanolamine-HCl (pH 7.5) were incubated for 15 min at 37 °C. Then, 0.1 vol of 10 $\times$ exonuclease III buffer (500 mM Tris-HCl (pH 7.5), 50 mM MgCl₂) and either 16 units exonuclease III or 250 units DNase I were added, and incubation was continued at 37 °C. At various times, aliquots (30 μ l) were taken into 10 μ l 0.5 M EDTA. The extent of

digestion was determined following precipitation with perchloric acid by measuring the release of acid-soluble ^{32}P counts.

Received 14 January; accepted 1 March 1999.

1. Friedberg, E. C., Walker, G. C. & Siede, W. *DNA Repair and Mutagenesis* (American Society for Microbiology, Washington, 1995).
2. Kanaar, R., Hoeijmakers, J. H. J. & van Gent, D. C. Molecular mechanisms of DNA double-strand break repair. *Trends Cell Biol.* **8**, 483–489 (1998).
3. Liang, F., Han, M. G., Romanienko, P. J. & Jasin, M. Homology-directed repair is a major double-strand break repair pathway in mammalian cells. *Proc. Natl Acad. Sci. USA* **95**, 5172–5177 (1998).
4. Takata, M. *et al.* Homologous recombination and non-homologous end-joining pathways of DNA double-strand break repair have overlapping roles in the maintenance of chromosomal integrity in vertebrate cells. *EMBO J.* **17**, 5497–5508 (1998).
5. Critchlow, S. E. & Jackson, S. P. DNA end-joining: from yeast to man. *Trends Biochem. Sci.* **23**, 394–398 (1998).
6. Sun, H., Treco, D. & Szostak, J. W. Extensive 3'-overhanging, single-stranded DNA associated with the meiosis-specific double-strand breaks at the *ARG4* recombination initiation site. *Cell* **64**, 1155–1162 (1991).
7. Sugawara, N. & Haber, J. E. Characterization of double-strand break-induced recombination: homology requirements and single-stranded DNA formation. *Mol. Cell. Biol.* **12**, 563–575 (1992).
8. Shen, Z. *et al.* Self-association of human Rad52 protein. *Mutat. Res.* **364**, 81–89 (1996).
9. Shinohara, A., Shinohara, M., Ohta, T., Matsuda, S. & Ogawa, T. Rad52 forms ring structures and cooperates with RPA in single-strand DNA annealing. *Genes to Cells* **3**, 145–156 (1998).
10. Van Dyck, E., Hajibagheri, N. M. A., Stasiak, A. & West, S. C. Visualization of human Rad52 protein and its complexes with hRad51 and DNA. *J. Mol. Biol.* **284**, 1027–1038 (1998).
11. Benson, F. E., Baumann, P. & West, S. C. Synergistic actions of Rad51 and Rad52 in genetic recombination and DNA repair. *Nature* **391**, 401–404 (1998).
12. Baumann, P., Benson, F. E. & West, S. C. Human Rad51 protein promotes ATP-dependent homologous pairing and strand transfer reactions in vitro. *Cell* **87**, 757–766 (1996).
13. Nussenzweig, A. *et al.* Requirement for Ku80 in growth and immunoglobulin V(D)J recombination. *Nature* **382**, 551–555 (1996).
14. Nussenzweig, A., Sokol, K., Burgman, P., Li, L. G. & Li, G. C. Hypersensitivity of Ku80-deficient cell lines and mice to DNA damage: the effects of ionizing radiation on growth, survival, and development. *Proc. Natl Acad. Sci. USA* **94**, 13588–13593 (1997).
15. Anderson, C. W. & Carter, T. H. in *Molecular Analysis of DNA Rearrangements in the Immune System* (eds Jessberger, R. & Lieber, M. R.) 91–112 (Springer, Heidelberg, 1996).
16. Rijkers, T. *et al.* Targeted inactivation of *MmRAD52* reduces homologous recombination but not resistance to ionizing radiation. *Mol. Cell. Biol.* **18**, 6423–6429 (1998).
17. Yamaguchi-Iwai, Y. *et al.* Homologous recombination, but not DNA repair, is reduced in vertebrate cells deficient in *RAD52*. *Mol. Cell Biol.* **18**, 6430–6435 (1998).
18. Game, J. C. DNA double-strand breaks and the *RAD50-RAD57* genes in *Saccharomyces*. *Semin. Cancer Biol.* **4**, 73–83 (1993).
19. Petes, T. D., Malone, R. E. & Symington, L. S. in *The Molecular and Cellular Biology of the Yeast Saccharomyces: Genome Dynamics, Protein Synthesis and Energetics* 407–521 (Cold Spring Harbor Laboratory Press, New York, 1991).
20. Mortensen, U. H., Bendixen, C., Sunjevaric, I. & Rothstein, R. DNA strand annealing is promoted by the yeast Rad52 protein. *Proc. Natl Acad. Sci. USA* **93**, 10729–10734 (1996).
21. Reddy, G., Golub, E. I. & Radding, C. M. Human Rad52 protein promotes single-strand DNA annealing followed by branch migration. *Mutat. Res.* **377**, 53–59 (1997).
22. Sugiyama, T., New, J. H. & Kowalczykowski, S. C. DNA annealing by Rad52 protein is stimulated by specific interaction with the complex of replication protein A and single-stranded DNA. *Proc. Natl Acad. Sci. USA* **95**, 6049–6054 (1998).
23. New, J. H., Sugiyama, T., Zaitseva, E. & Kowalczykowski, S. C. Rad52 protein stimulates DNA strand exchange by Rad51 and replication protein-A. *Nature* **391**, 407–410 (1998).
24. Shinohara, A. & Ogawa, T. Stimulation by Rad52 of yeast Rad51-mediated recombination. *Nature* **391**, 404–407 (1998).
25. Sung, P. Function of yeast Rad52 protein as a mediator between replication protein-A and the Rad51 recombinase. *J. Biol. Chem.* **272**, 28194–28197 (1997).
26. Gasior, S. L., Wong, A. K., Kora, Y., Shinohara, A. & Bishop, D. K. Rad52 associates with RP-A and functions with Rad55 and Rad57 to assemble meiotic recombination complexes. *Genes Dev.* **12**, 2208–2221 (1998).
27. Sung, P. Yeast Rad55 and Rad57 proteins form a heterodimer that functions with replication protein-A to promote DNA strand exchange by Rad51 recombinase. *Genes Dev.* **11**, 1111–1121 (1997).
28. Subramani, S. & Seaton, B. L. in *Genetic Recombination* (eds Kucherlapati, R. & Smith, G. R.) 549–573 (American Society for Microbiology, Washington, DC, 1988).
29. Johnson, B. L., Thyagarajan, B., Krueger, L., Hirsch, B. & Campbell, C. Elevated levels of recombinational DNA repair in human somatic cells expressing the *Saccharomyces cerevisiae RAD52* gene. *Mutat. Res. DNA Repair* **363**, 179–189 (1996).
30. Park, M. S. Expression of human *RAD52* confers resistance to ionizing radiation in mammalian cells. *J. Biol. Chem.* **270**, 15467–15470 (1995).

Acknowledgements. We thank our colleagues for their interest, suggestions and careful reading of the manuscript. We also thank J. Dubochet for his interest in the project, and N. Hajibagheri for related electron microscopic analyses. This work was supported by the Imperial Cancer Research Fund, the Human Frontiers Science Program, the Swiss National Foundation and the Swiss-British Council Joint Research Program. E.V.D. was supported in part by an EC fellowship.

Correspondence and requests for materials should be addressed to S.C.W. (e-mail: s.west@icrf.icnet.uk).